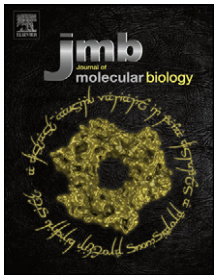
Review



New Biological Insights from Better Structure Models

Wouter G. Touw[1], **Robbie P. Joosten**[2] **and Gert Vriend**[1]

*1 - **Centre for Molecular and Biomolecular Informatics,** Radboud University Medical Center, Geert Grooteplein-Zuid 26-28, 6525 GA Nijmegen, The Netherlands*
*2 - **Department of Biochemistry,** Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*

Correspondence to Gert Vriend: *Gerrit.Vriend@radboudumc.nl*
http://dx.doi.org/10.1016/j.jmb.2016.02.002
Edited by M. Sternberg

Abstract

Structure validation is a key component of all steps in the structure determination process, from structure building, refinement, deposition, and evaluation all the way to post-deposition optimisation of structures in the Protein Data Bank (PDB) by re-refinement and re-building. Today, many aspects of protein structures are understood better than 10 years ago, and combined with improved software and more computing power, the automated PDB_REDO procedure can significantly improve about 85% of all X-ray structures ever deposited in the PDB. We review structure validation, structure improvement, and a series of validation resources and facilities that give access to improved PDB files and to reports on the quality of the original and the improved structures. Post-deposition optimisation generally leads to improved protein structures and a series of examples will illustrate how that, in turn, leads to improved or even novel biological insights.

Introduction

In 1951, Pauling and Corey predicted the α-helix [1] and the β-sheet [2]. In 1958, the first picture of a protein was obtained when Kendrew and colleagues solved the structure of myoglobin at 6 Å using X-ray crystallography [3]. In 1960, the structure of myoglobin was obtained at 2 Å [4] and the structure of haemoglobin was solved [5]. The similarity between the tertiary structures of haemoglobin and myoglobin showed the evolutionary conservation of the globin folds [6,7], and these structures "laid the foundation" [8] for understanding the mechanism of cooperativity in haemoglobin and hinted already at the possibility to perform homology modelling, which was performed for the first time when α-lactalbumin was modelled based on the crystal structure of hen egg-white lysozyme 4 years later [9]. Myoglobin and haemoglobin illustrate the impact of protein atom coordinates on science in general and on biology in particular. Kendrew and Perutz received for their work a Nobel Prize, an honour later also bestowed on scientists[†] for (structure) work on G protein-coupled receptors, the ribosome, insulin, the photosynthetic reaction centre, ATPase, GFP, ubiquitin, ion and water channels, protein structure NMR in general, and for computational techniques on protein structures. The last of this list was awarded for "the development of multiscale models for complex chemical systems", which is a computational technique that critically depends on the accuracy of the protein structure coordinates.

Novotný, Bruccoleri, and Karplus were the first to ask whether correctly folded protein models could be distinguished from incorrectly folded protein models [10,11]. They modelled the sequence of the α-helical sea worm haemerythrin on the mainly β-stranded mouse immunoglobulin VL domain and vice versa. The incorrect side chains could be incorporated reasonably well and the empirical potential energy of these mis-folded models was comparable to the correct models. The mis-folded models, however, had higher non-covalent energy terms, a larger solvent-accessible surface area, and more exposed non-polar side-chain atoms [10]. Others reported that, compared to the correct models, the incorrect models have a lower solvation free energy [12], are less compact [13], and

make only about half as many hydrophobic contacts [14]. These deliberately mis-folded proteins have long served to validate protein structure analysis and validation methods [12–24].

Validation

It is generally believed that the field of protein structure validation came into existence in 1989 when serious errors were discovered in a series of deposited crystal structures (see Fig. 1). These discoveries led to the Collaborative Computational Project Number 4 study weekend "Accuracy and reliability of macromolecular crystal structures" and Brändén and Jones subsequently published their seminal commentary on errors and checks to detect them [39]. Soon after this commentary was published, the crystallographic model-building program O could compare rotamers and the position of backbone oxygen atoms to database penta-peptides [40]. The protein structure bioinformatics community also took up the challenge

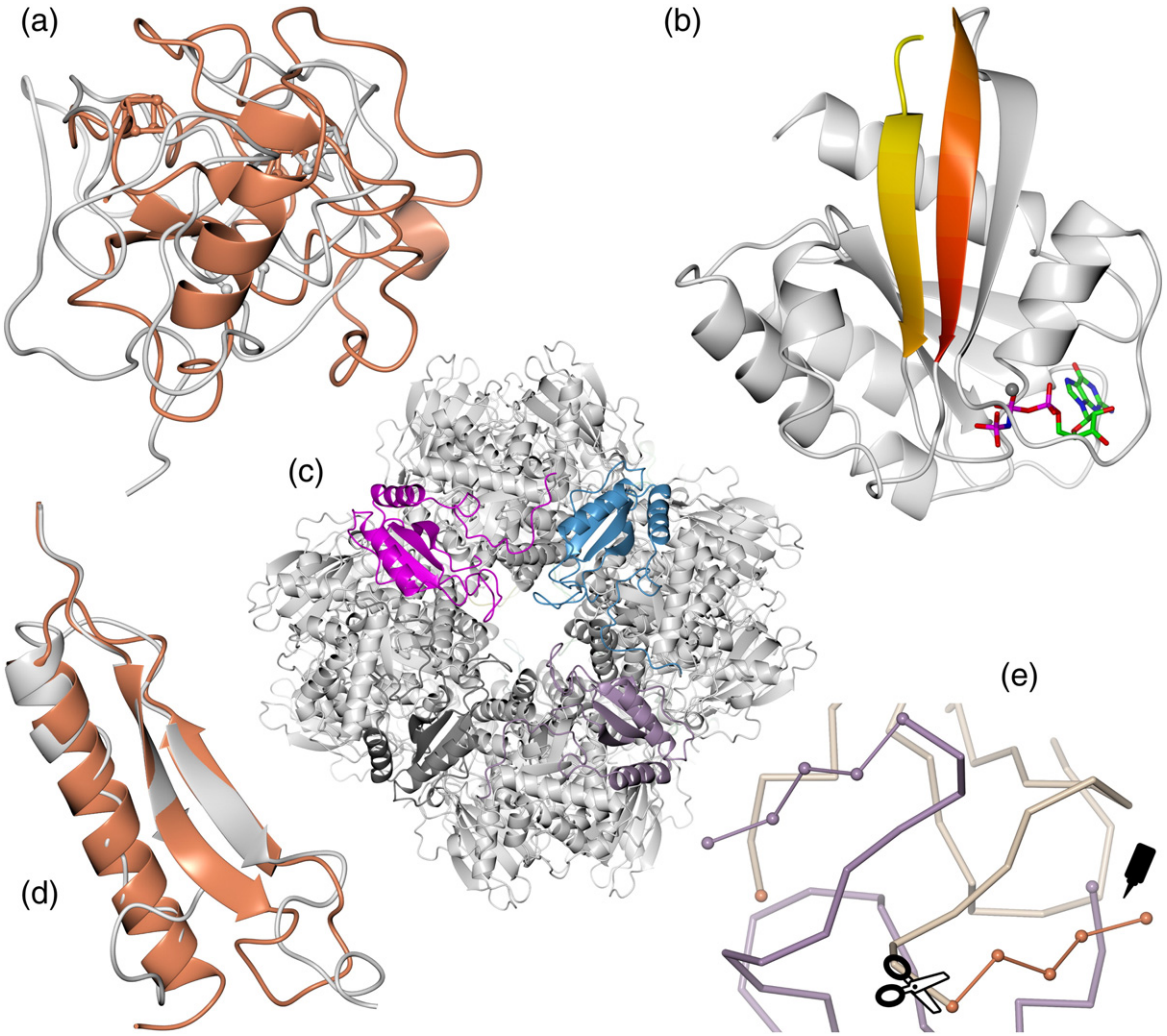


Fig. 1. Threading errors in protein structures. (a) The best superposition possible for ferredoxin I in the correctly traced PDB entry 5fd1 [25] (orange) and the mis-threaded [26] PDB entry 2fd1 [27] (grey). (b) β-Strands β1 (yellow) and β3 (orange) in the correct PDB entry 5p21 [28]. These two β-strands were traced in each other's density [29] in the structure of human p21 reported in Ref. [30]. (c) The small subunits of spinach RuBisCo are coloured in the correct structure model 1rcx [31]. The small RuBisCo subunit was threaded incorrectly [32] in the structure reported in Ref. [33]. (d) Residues 143–203 of the enolase structures 1enl [34] (grey) and 2enl [35] (orange). The first β-strand is followed by a loop and an anti-parallel second β-strand in the correct structure model 2enl. The 1enl structure model is traced backwards. (e) In the C^{α}-only PDB entry 2hvp [36], five C^{α} atoms (ball-and-sticks) are incorrectly assigned to the C-terminus of the HIV-1 protease rather than correctly to the N-terminus (spheres), resulting in an erroneous dimer interface [37]. The problem can be resolved (PDB entry 3hvp [37]) by breaking the C-terminal connection (scissors) with the stretch of five C^{α} atoms and connecting (glue) the stretch to the N-terminus instead. The symmetry-related copy is shown in purple. Figures were prepared with CCP4mg [38].

and, in 1993, the first three structure validation methods were published in rapid succession: Directional Atomic Contact Analysis (DACA [21]), PROCHECK [41], and Protein Structure Analysis (ProSA [23]) were published. These three methods determined rules from protein structures solved at high resolution—that are therefore presumed "correct"—to find errors in protein structures in general.

Vriend and Sander determined a contact quality index that measures the agreement between the atom distributions of all possible close contacts in the structure model and equivalent database distributions [21]. This detailed evaluation of atomic packing also allows for the detection of local errors in the protein packing. This method was implemented in WHAT IF [42]. The quality index resulting from this analysis is now known as packing quality or DACA in WHAT_CHECK [43].

Thornton and co-workers described the stereochemical quality of protein structures in terms of the parameters derived by Morris *et al*. (C^{α} chirality, disulfide bond length, proline ϕ, main-chain hydrogen bond energy, peptide bond planarity, side-chain torsion angles χ1 and χ2, etc.) [44], bond lengths and bond angles [45], and position in the Ramachandran plot [46]. These stereochemical checks were implemented in PROCHECK [41].

Sippl applied the concept of potentials of mean force [47] to C^{α}–C^{α} [23] and C^{β}–C^{β} [16,23] distances. In the ProSA method, the pseudo-energy of proteins is derived using a combination of mean force potentials, and the mean field energy of a protein structure is transformed into a *Z*-score by evaluating the energy for a large number of structure decoys (alternative conformations) as well [23].

More methods that give a score to a whole molecule have been published; often, like ProSA, in threading projects. Eisenberg and colleagues, for example, calculated amino acid preferences as a function of three-residue environment parameters (the area of the residue that is buried, the fraction of side-chain area that is covered by polar atoms, and the local secondary structure) [48] and measured the compatibility of a protein model with its sequence using this so-called three-dimensional profile [19].

Initially, protein structure validation was met with some resistance from the protein structure determination field (e.g., see Ref. [53]), but some high-profile cases of structure models with very unusual features (see Fig. 2) led the protein structure communities—structure determination and bioinformatics alike—to start Validation Task Forces (VTFs) for X-ray crystallography [54], NMR [55], and electron microscopy [56]. The X-ray and NMR VTFs have written their recommendations, and the wwPDB consortium [57] is presently implementing these recommendations in software that depositors of structures must use. It will take time to implement all VTF recommendations; thus, depositors who want to very extensively validate their structure before deposition will still need, for a while, to use tools such as WHAT_CHECK[‡], MolProbity [58],

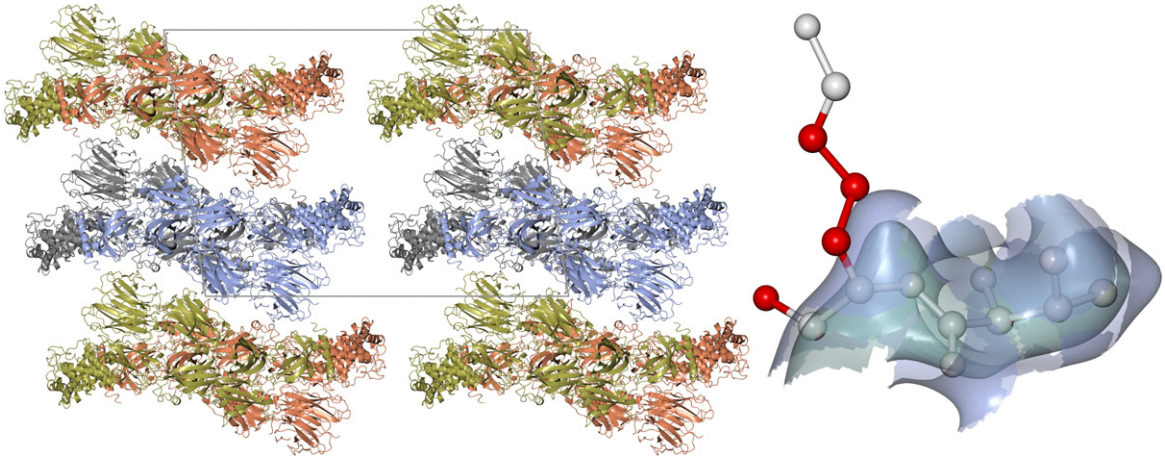


Fig. 2. Protein structures with improbable features. Left: in 2006, Gros and colleagues [49] identified several unusual and improbable features in a structure of the complement protein C3b (PDB entry 2hr0). The absence of crystal contacts in the c-direction of the unit cell (grey) was the most improbable feature. This triggered an investigation by the University of Alabama as to whether structures solved by H. K. M. Murthy were fabricated [50]. Right: another instance of unusual features was discovered by Rupp in 2012 [51]. The figure shows residues Val134 and Lys135 (red: atoms modelled at zero occupancy; white: full occupancy) of PDB entry 3k78 [52] and the $2mF_o - DF_c$ map (calculated using a grid size of 0.1 Å) contoured at +1.0 σ (green) and at the noise level +0.4 σ (blue). One of the highly improbably model features noticed by Rupp was the complete absence of any $2mF_o - DF_c$ density for unoccupied atoms down to near-noise levels whilst normal main-chain *B*-factors had been reported. This suggested that the data were indeed calculated from a model with zero occupancy atoms [51].

and CING [59] in addition to the Protein Data Bank (PDB) [57,60] validation server [54,61].

Validation tools can be categorised in many ways; for example, by their level of detail. Most of the older tools give one score for the whole structure and capture the overall quality of a structure in one number. Although ProSA, PROCHECK, DACA, and QMEAN [62,63], for example, score aspects of individual residues, their strength lies in whole protein quality evaluation. Validation tools also can be categorised by the certainty with which they can call things right or wrong. Most validation options do not determine the quality but merely the normality of a protein structure, that is, how similar a protein structure model is to a collection of good, high--resolution structures in terms of the validation parameters. Other validation options (noticeably, all nomenclature checks, many administrative validation options, and routines that calculate the agreement of a model with the experimental data) provide answers about quality rather than normality. The Cambridge Structural Database (CSD) [64] holds more than 700,000 structures of small molecules that have been solved at much higher resolution than most PDB entries. Geometric parameters that can be determined from an analysis of CSD files are therefore so accurate that they can be used for all practical purposes as a gold standard when solving or validating PDB entries. The prime example of CSD-derived parameters are the famous Engh and Huber bond length and bond angle data [45,65] that are still used today in most refinement and validation software. Similarly, Hooft *et al*. used the CSD to determine the normal deviation from planarity in planar groups in proteins [66].

Global scores can detect bad structure models but they are not very useful when validation is used to actually improve model quality. Many tools, fortunately, detect erroneous molecular details that can be used directly to improve the quality of structures.

Hooft *et al*. used the CSD to arrive at a force field for hydrogen bond energies and used this force field to optimise the flipping of Asn, Gln, and His side chains [67]. Hydrogen bond network optimisation is part of WHAT_CHECK [43] and MolProbity [68]. Nielsen showed the importance of this validation-based structure improvement for electrostatic calculations [69,70], and the realisation that a series of measured pK_a values commonly used for the calibration of electrostatic computation methods were flawed by crystal packing artefacts dramatically improved the entire field of protein electrostatics [70]. Nielsen also showed that electrostatic calculations for most enzymes in the PDB would give significantly better results if the hydrogen bonding network would be improved prior to the calculations [69].

Wrong cell dimensions lead to systematic deviations in bond lengths and angles. Hooft *et al*. wrote software that projects the protein's bond lengths and angles on the axis system of the crystal cell to correct the cell's dimensions [71]. Lamzin and co-workers later improved this method [72].

The growth of the PDB has allowed validation of the Ramachandran plot [58,73] and bond lengths and angles [74,75] to become specific for secondary structure and residue type. Rotamer libraries constructed using high-resolution protein structures [40,76–79] are used in several programs to perform a knowledge-based validation of side-chain conformation. Misfit side chains may also be detected by C^{β} position deviations [80] and steric clashes [43,58]. The RosettaHoles software provides a validation score for under-packing [81].

Several groups have developed tools such as VHELIBS [82], ValLigURL [83], Mogul [84], and Twilight [85] to visualise and validate ligands; tools such as pdb-care [86], CARP [87], and Privateer [88,89] to check carbohydrates [90,91]; and programs such as ERRASER [92] to re-build nucleic acids and MolProbity [58] to validate nucleic acids.

CH_4, NH_3, NH_4^+, H_2O, OH^-, Ne, Na^+, Mg^{2+}, and Al^{3+} all contain 10 electrons and thus will scatter X-rays roughly equally much. This makes it hard to see the difference between them in any electron density determined at worse than atomic resolution. The same problem exists for K^+ and Ca^{2+} that both have 18 electrons. Moreover, K^+ and Ca^{2+} at half-occupancy scatter X-rays roughly equally much as H_2O, Na^+, Mg^{2+}, and so on. Consequently, many ions in the PDB are of the wrong type or actually should be water, whilst many waters should be ions [93]. Brown has determined empirical bond valence parameters [94,95] that can be used to determine the ion type from the distances between the ion and its coordinating atoms. This method works reasonably well but only at high resolution when all surrounding atoms can be seen very well in the density and when there is no bias caused by refining ion X as ion Y forcing ion X to get the ligand-atom distances of ion Y. The Brown parameters have been implemented in SHELX [96] and WHAT_CHECK (unpublished results) and, later also, in CheckMyMetal [97] and Phenix [98].

Alkali and alkaline earth metals are preferentially coordinated by oxygen atoms and not by nitrogen atoms. Dauter *et al*. recently reported that calcium ions in the so-called strong calcium site of several savinase structure models seem to be coordinated by the nitrogen atom instead of the oxygen atom of an asparagine [99]. A pseudo-octahedral calcium site is normally coordinated by oxygen atoms only. The *B*-factors of the Asn suggested that the side chain should be flipped. An inspection of the PDBREPORT database [43] reveals that the chemically highly implausible coordination of sodium, potassium, calcium, or magnesium ions by the nitrogen atom instead of the oxygen atom in asparagine or glutamine side chains occurs in 327 sites in 269 PDB entries,

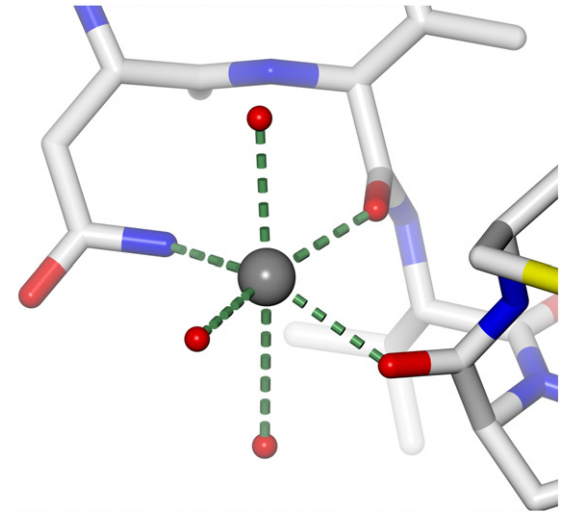


Fig. 3. Asn side-chain flips at atomic resolution. Sodium-binding site of glycinamide ribonucleotide transformylase (PDB entry 1kjq, determined at 1.05 Å resolution [100]). The side chain of Asn100 has a highly unlikely conformation because the nitrogen rather than the oxygen coordinates the ion. PDB_REDO flips the side chain so that the ion is coordinated by the side-chain oxygen. The PDB files of several hundred incorrect sites similar to this site contain LINK records specifying the N-metal coordination. Incorrect LINK records between the nitrogen of Asn and Gln side chains and Na^{+}, Mg^{2+}, K^{+}, and Ca^{2+} are removed, allowing structure correction by side-chain re-building (since PDB_REDO version 5.37). Future versions of PDB_REDO will also flip metal coordinating side chains based on results from WHAT_CHECK.

including the savinase example and several structures solved at atomic resolution (Fig. 3).

Several authors have noted an under-representation of *cis* peptides in the PDB that is partly the result of the a priori assumption in structure determination that all peptides have a *trans* conformation [101–104]. *Trans* peptide planes have also been observed rotated by 180°; this is called a peptide plane flip. Peptide plane flips typically are the result of mistakes in the early stages of model building when the electron density maps are not yet very clear. An incorrectly built peptide plane tends to lead to locally distorted geometry. We recently designed a random-forest-based method to detect these problems [105] and found almost 5000 *trans–cis* errors and many thousands of peptide plane flips.

Improving PDB files by re-refinement and re-building

Over the years, 3299 PDB entries have been made obsolete. Most times, these entries were made obsolete because a better version became available, for example, based on higher-resolution data, but sometimes, the entries were highly improbable and made obsolete without putting a replacement file in the PDB. We also see more and more cases of PDB files that were improved and deposited by others than the original authors. In 2007, Joosten and Vriend took a more systematic approach and re-refined some 1200 structure models for which data were available to 2.00 Å resolution [106]. More than three-quarters of the re-refined models had an improved *R*-free value and improved geometric characteristics. After this successful small-scale proof of concept, Joosten *et al.* re-refined all high- and medium-resolution X-ray structures in the PDB (15,000 at the time) for which the reflection data (including the *R*-free set) were deposited and useful [107]. They showed that, despite the complication of using many more low--resolution models, two-thirds of the re-refined models were improved in terms of *R*-free [107]. The Ramachandran *Z*-score of the structure models also improved over the entire resolution range. In addition, they showed that the possibility of improving published structure models was not limited to old structure models but that more than 60% of recently deposited structure models could also be improved. The addition of side-chain re-building and peptide--flipping tools [108] plus more advanced refinement parameterisation algorithms improved the success rate of PDB_REDO and extended the scope to active correction of modelling errors or, more poetically, “constructive validation” [109]. The specific PDB_REDO steps that are applied based on validation algorithms are described in Table 1. Analyses of *R*-free and six WHAT_CHECK model-quality metrics of 12,000 randomly chosen PDB entries show that 85% of the PDB entries can be improved in terms of overall quality [109].

The re-refined structure models (plus electron density maps and a multitude of metadata) are stored in the PDB_REDO databank [112,113]. This repository now holds 99% of all crystallographic PDB entries for which experimental data are deposited (currently more than 90,000). New entries are added automatically with every new PDB release. Older PDB_REDO entries are replaced gradually or whenever a PDB entry is re-released, typically because of changes in the entry's annotation. It should be noted that many changes in annotation of PDB entries are the result of the PDB_REDO project. Over the course of the project, nearly 7500 annotation problems that somehow hampered the optimisation or interpretation of PDB entries were reported and the PDB staff corrected the majority of these.

In an automated procedure, there is always a risk of introducing errors. A particularly difficult step is the restraint generation for ligands. This relies on reasonable input coordinates of a ligand, correct annotation of the chemistry by the PDB and/or

Table 1. Validation-driven PDB_REDO steps

PDB_REDO step	Programs involved
Removal of improbable (metal coordination) LINKs	Stripper [109]
Correction of carbohydrate LINK topology	Stripper
Correction of carbohydrate names	pdb-care [86] and stripper
Removal of superfluous carbohydrate oxygens	pdb-care and stripper
Removal of improbable ligand occupancy models	REFMAC [110]
Removal of overly detailed *B*-factor models	REFMAC and Bselect [109]
Correction of atomic chirality problems	REFMAC, WHAT_CHECK [43], and Chiron [109]
Addition of missing side-chain atoms	SideAide [108] and DSSP [111,112]
Histidine, asparagine, and glutamine flips to improve hydrogen bonding	WHAT_CHECK and SideAide
Peptide flipping	Pepflip [108] and DSSP
Removal of waters not supported by the electron density	Centrifuge [108]

proper interpretation of the coordinates by the tools in PDB_REDO. Although ligands generally improve slightly in PDB_REDO [82], sometimes ligands are refined incorrectly. It is therefore highly recommended to critically inspect ligands and their electron density manually, which is, by the way, not different for original PDB entries [83,85,114,115].

Taken together, the PDB_REDO procedure typically leads to structure models that better fit their experimental data, have more plausible molecular geometry, and are more informative for biological interpretation.

Better biology through better structure models

Four decades after Browne's first attempt at homology modelling on α-lactalbumin, the technique has become a research field in itself, and much effort is directed towards selecting good models from a large set of candidates (e.g., see Ref. [116] and references therein). Using both PDB templates and PDB_REDO templates, we built homology models with YASARA for 33 CASP11 targets for which the alignment is essentially certain whilst small structural details are important (best GDT_TS > 60%) [116]. We found that the average C^{α} root-mean-square (RMS) error was reduced from 2.28 Å (using PDB templates) to 2.15 Å (using PDB_REDO templates). The average C^{α} RMS deviation between PDB and PDB_REDO structure models is 0.15 Å. These results suggest that the use of PDB_REDO templates certainly does not harm the homology modelling process and that the changes made by PDB_REDO are improvements in the right direction.

Better template structures thus lead to better homology models, and both better structure models and better homology models obviously must lead to better answers to biological questions. Sometimes corrections in PDB files do not influence the biology; a bond length correction by 0.18 Å, for example, is crystallographically significant but will not change the answer to a question related to mutability, antigen selection, or intermolecular interactions. Other corrections though, for example, replacing a calcium ion near the active site by a zinc or flipping the side chain of asparagine in the ligand binding pocket, are likely to lead to radically different and more reliable answers to questions related to understanding an enzyme's mechanism or drug design.

The next 10 sections review examples of “better biology through better structures”. In each example, improvement of the PDB file led to a different view on the biological role of a molecule or to a different answer to a biological question.

Example 1—Peptide plane flip in Plk1 Polo Box Domain substrate

García-Álvarez *et al.* reported the structure of the Polo Box Domain (PBD) of the human serine/threonine kinase Plk1 in complex with a 9-mer phosphopeptide substrate derived from Cdc25C (PDB entry 2ojs [117]). Plk1 is essential for regulating cell cycle progression and is an important drug target for cancer therapy [118]. They discuss the molecular mechanisms of substrate recognition of Plk1 and the implications for the centrosomal localisation and activity.

The PDB_REDO program *Pepflip* detected that the peptide plane between Leu1 and Leu2 of the Cdc25C phosphopeptide should be flipped to better fit the electron density and improve the Ramachandran plot. In the correct conformation, there is an additional hydrogen bond between the peptide and Asp416 in Plk1 (Fig. 4). Furthermore, after the peptide plane flip, the Cdc25C peptide forms an additional β-strand, thereby extending the β-sheet in Plk1. The substrate conformation was corrected by the depositors who obsoleted PDB entry 2ojs and superseded it by PDB entry 3bzi. Free energy calculations using the corrected phosphopeptide showed that the phosphothreonine residue and the main-chain atoms of the peptide account for the majority of the binding enthalpy [119].

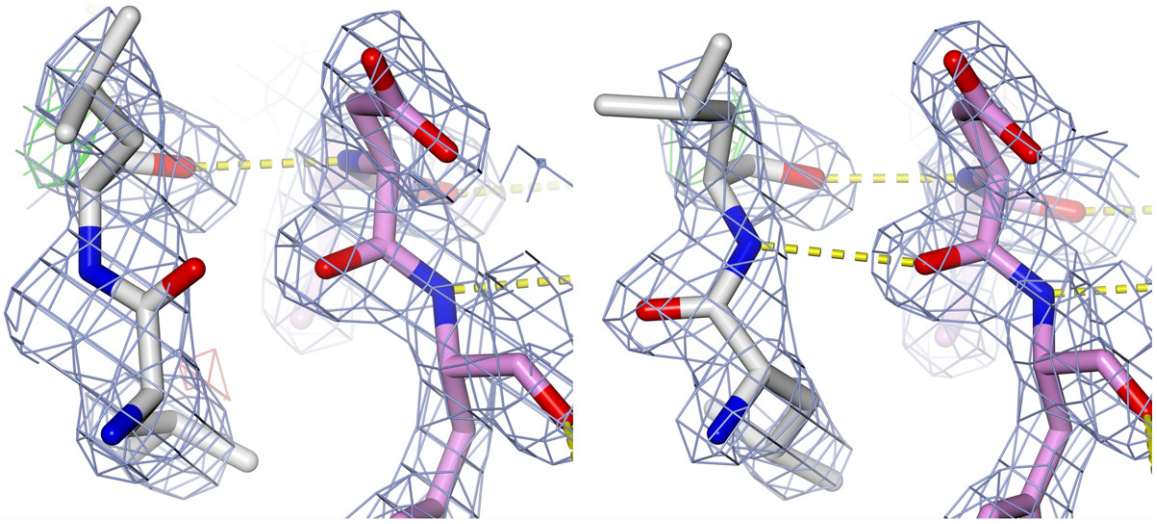


Fig. 4. Peptide plane flip in Plk1 PBD substrate. Left: the electron density around the peptide plane between Leu1 and Leu2 of the substrate (white carbons) suggests that the peptide should be flipped to allow a hydrogen bond to the carbonyl of Asp416 in the enzyme (pink carbons) in PDB entry 2ojs [117]. Right: the peptide substrate extends the Plk1 β-sheet in the superseding entry 3bzi due to the flipped peptide plane. Unless mentioned otherwise, the $2mF_o - DF_c$ and $mF_o - DF_c$ maps have been sampled with a grid size equal to a third of the resolution and are shown at a contour level of +1.2 σ (blue) and +3 σ (green) and −3 σ (red), respectively, and the entire model has been used for the calculated structure factors. The $2mF_o - DF_c$ and $mF_o - DF_c$ maps are shown up to 2 Å from the displayed peptide atoms.

The extension of the β-sheet is an integral part of the substrate recognition mechanism and is only visible in the corrected structure.

Example 2—Wishfully modelling a Plk1 PBD inhibitor

The Plk1 PBD binds to pThr/pSer-containing motifs [120]. Qian *et al*. reported a Plk1 PBD structure in complex with an inhibitor in the PBD pocket (PDB entry 4mlu [121]). Qian *et al*. designed the inhibitory peptide to mimic a natural substrate but wanted to improve the cellular uptake efficiency by making the inhibitory peptide mono-anionic rather than di-anionic by masking the phosphothreonine. The mono-anionic phosphoester was fitted in the reported structure model 4mlu.

Dauter *et al*. discovered that the electron density does not justify modelling the phosphoester moiety (Fig. 5) [99]. Thus, the inhibitor is still di-anionic. Qian *et al*. then retracted their paper and replaced the phosphoester moiety by water in PDB structure 4o6w that supersedes 4mlu (Fig. 5).

In summary, the design of PBD inhibitors that both mimic the natural substrate and have drug-like physicochemical properties is still an open challenge.

Example 3—hPHMT ligand identification

Human phenylethanolamine *N*-methyltransferase (hPNMT) catalyses the conversion of *R*-noradrenaline into *R*-adrenaline. In this reaction, a methyl group is transferred from the co-factor *S*-adenosyl-L-methionine to noradrenaline. PNMT inhibitors specific to the central nervous systempotentially are important drug targets for Alzheimer's and Parkinson's diseases. In a fragment-based drug design screen, Drinkwater *et al*. soaked hPNMT crystals with 96 mixtures of four chemically diverse small molecules and modelled 12 hits in the electron density, 9 of which were confirmed by isothermal titration calorimetry to bind to hPNMT [122].

Nair *et al*. showed that molecular dynamics (MD) simulations reproduced the crystal structure binding mode modelled for these nine compounds [123]. For one of the other cocktails, Drinkwater *et al*. proposed 6-chlorooxindole as the most likely candidate for explaining the electron density observed in the noradrenaline pocket (PDB entry 3kpy, Fig. 6), but they could not confirm binding by isothermal titration calorimetry. The MD simulations predicted that 6-chlorooxindole cannot stably bind to hPNMT. In contrast, the simulations suggested that the pocket was occupied by two other fragments in the cocktail, benzene-1,3-diol and imidazole. Free energy calculations predicted the binding to be cooperative and re-refinement showed that these two fragments together could also account for the electron density (PDB entry 4dm3, Fig. 6).

The combined pharmacophores of benzene-1,3--diol and imidazole provide a better basis for rational

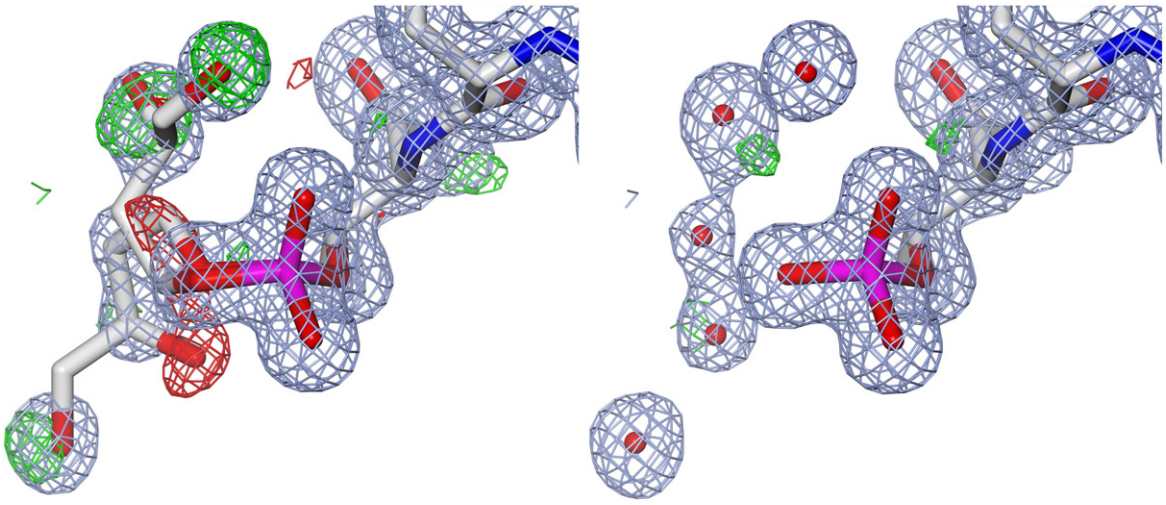


Fig. 5. The phosphothreonine fragment of a PBD inhibitor in the structure of the human Plk1 is di-anionic rather than mono-anionic. Left: although a disordered phosphoester moiety is modelled in PDB entry 4mlu [121], the electron density suggests that this group is absent. Right: water molecules and a di-anionic inhibitor are modelled in the superseding PDB entry 4o6w [121]. The 2mF_o – DF_c and mF_o – DF_c maps are shown up to 1.5 Å from the displayed inhibitor fragment and water molecules.

design and thus for the development of hPNMT inhibitors.

Example 4—Herceptin-HER2 interface

When over-expressed, human epidermal growth factor receptor 2 (HER2, also known as ErbB2 and Neu) can promote malignant cell transformation [124]. The monoclonal antibody Trastuzumab, commercially known as Herceptin, is known to have an anti-proliferative effect on cells transformed by over-expression of HER2 and is therefore used to treat HER2-positive metastatic breast cancers [125]. The structure of the Fab fragment of Herceptin bound to the extracellular domain of HER2, PDB entry 1n8z [126], shows the binding interface of the two proteins. This indicates where Herceptin binds, but as a result of poor side-chain fitting, the structure model does not properly show how and why Herceptin binds (Fig. 7).

Automated re-building of 1n8z reveals numerous additional receptor–antibody interactions, resulting in a much more faithful description of the binding mode of Herceptin.

The better understanding of the binding mode of Herceptin contributes to the development of other monoclonal antibodies in cancer immunotherapy.

Example 5—Ion identity in myosin heavy chain kinase regulatory sites

Myosin II plays a central role in cytokinesis, cell migration, and adhesion [127]. The α-kinase domain of myosin heavy chain kinase (A-CAT) is involved in regulating the formation of myosin II filaments and the active site of A-CAT undergoes a conformational switch that is said to be influenced by the magnesium-binding sites [128].

Minor and co-workers recently implemented Brown's bond valence method in the CheckMyMetal web server [97] for the validation of metals in macromolecular structures. They reported several examples of mis-identified ions, amongst which the magnesium ions in A-CAT (PDB entry 3lkm, [128]). The validation results, the reported crystallisation conditions, the sample preparation, and manual re-refinement all suggest that one magnesium ion should be replaced with water, whilst the other two should be replaced by potassium and coordinated also by ethylene glycol (Fig. 8) [97].

The presence of potassium rather than magnesium in the regulatory sites casts serious doubt on the role of magnesium and suggests that the role of potassium in regulating the activity of α-kinase is worth investigating.

Example 6—*Trans*–*Cis* isomerisation in Rab4a switch 2 region

The Ras-like protein Rab4 is involved in endosomal sorting by orchestrating a small GTPase cascade for recruitment of adaptor proteins to early endosomes [129]. Despite the high level of sequence similarity between members of the Rab family, each member targets specific effector proteins. One of the molecular regions involved in the discrimination between different effector

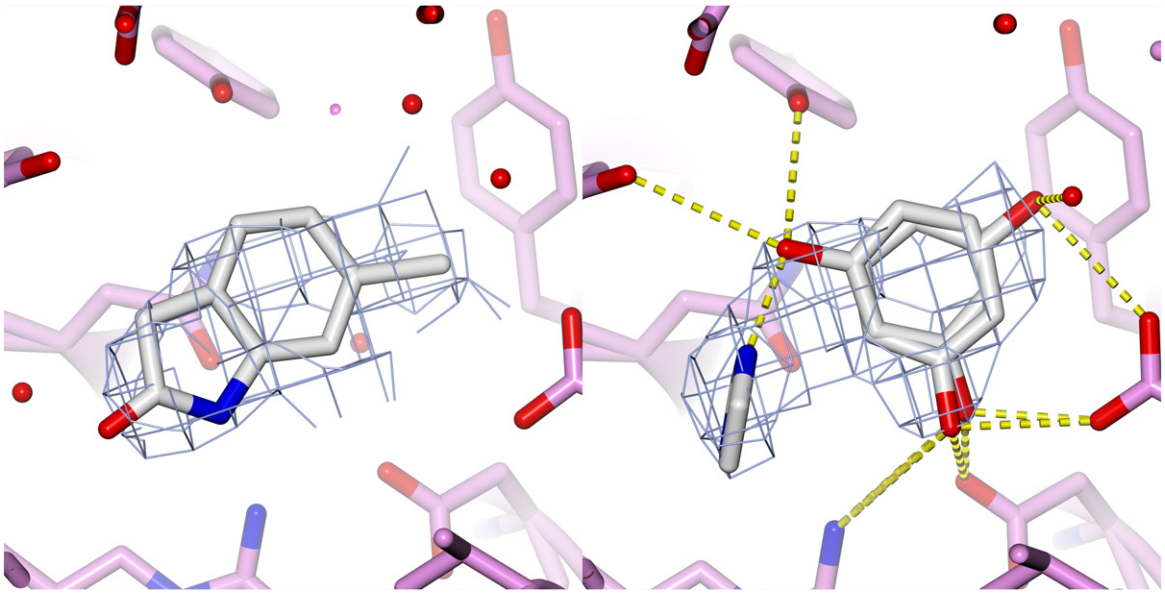


Fig. 6. Ligand identification in the hPNMT active site. The binding pocket is occupied by 6-chlorooxindole in PDB entry 3kpy [122] (left) and with benzene-1,3-diol (in two alternative conformations) and imidazole in PDB entry 4dm3 [123] (right). The binding modes of these two ligands were predicted by MD simulations [123]. The figure shows the possible hydrogen bonds over time. The 2mF_o – DF_c map is shown up to 1.5 Å from the ligands.

proteins is the so-called switch 2 region [130]. The switch 2 region is rearranged upon GTP hydrolysis. The structure of human Rab4a has been solved in the active state with the GTP analogue GppNHp (PDB entry 2bme [130]) and in the inactive GDP-bound state (PDB entry 2bmd [130]).

Residue Phe72 is located at the start of α-helix H2 in the switch 2 region of Rab4a and has the *trans* conformation in the GppNHp-bound state. The *trans* conformation is also present in the GDP-bound structure. Recently, a method was created to detect *cis* peptides erroneously modelled as *trans* peptides [105]. The method predicted that Phe72 in the GDP-bound state should have been modelled as a *cis* peptide rather than a *trans* peptide and this prediction was validated by re-refinement [105] (see Fig. 9).

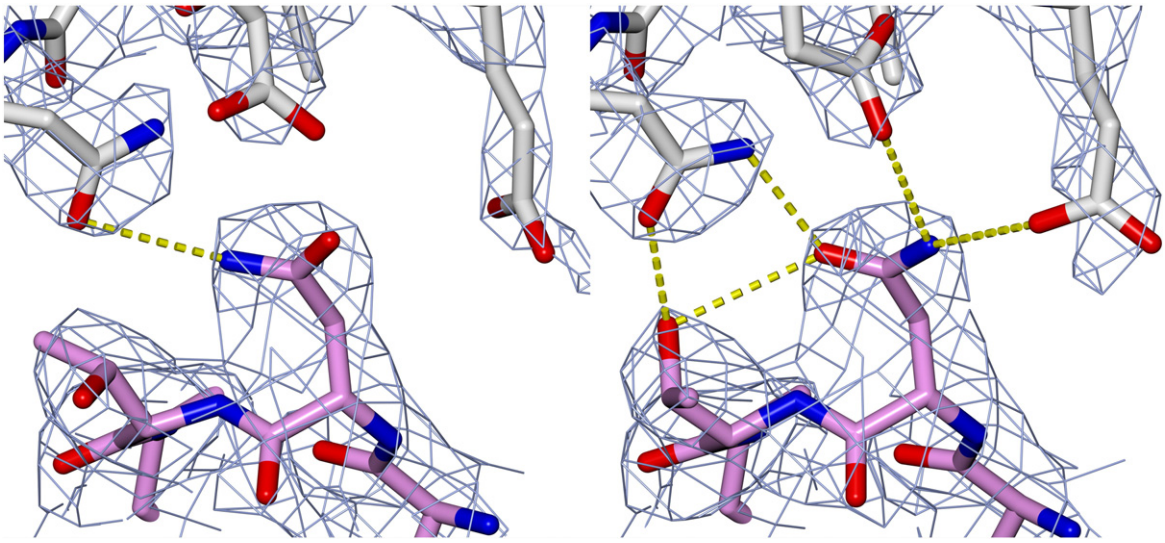


Fig. 7. Improving the binding interface between Herceptin and HER2. Left: detail of PDB entry 1n8z showing the Fab light chain of Herceptin (pink) with a single hydrogen bond to HER2 [126]. Right: the PDB_REDO optimised version of 1n8z. Flipping Asn30 and re-fitting Thr31 together with small adjustments to the local HER2 side chains reveal a hydrogen bonding network between the proteins containing four hydrogen bonds and one hydrogen bond that correctly positions the Thr31 and Asn30 side chains. The 2mF_o – DF_c map is shown up to 1.5 Å from the protein fragments.

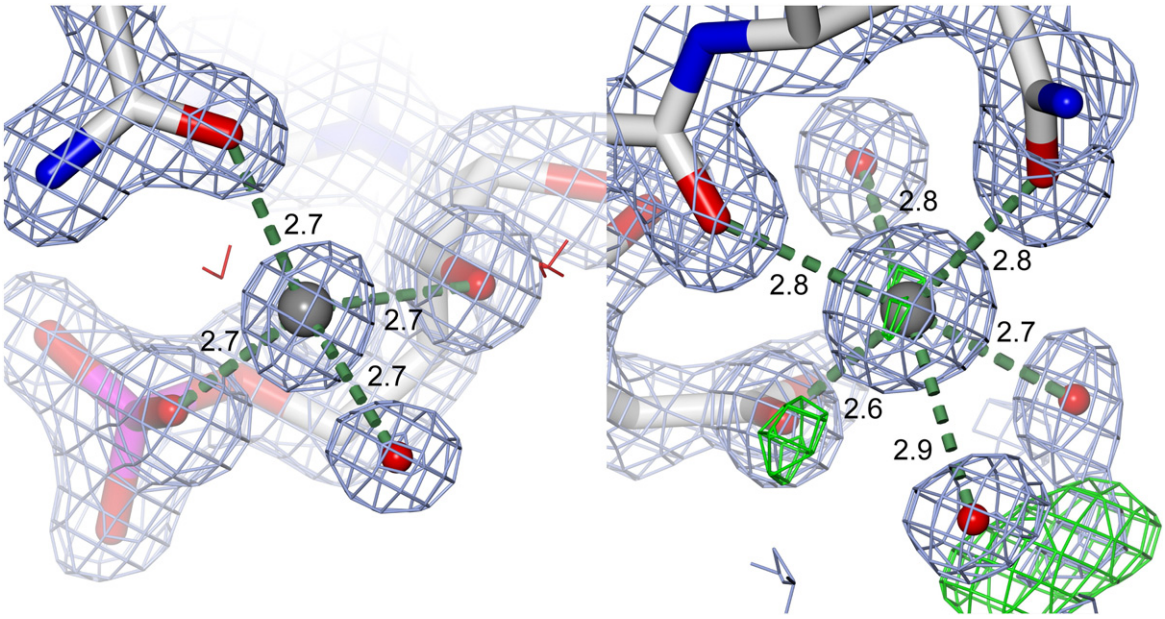


Fig. 8. Water and potassium rather than magnesium in the α-kinase domain of myosin heavy chain kinase. Left: Mg901 should be replaced by a water molecule in PDB entry 3lkm [128] according to metal validation software [97]. The contact distance is shown in angstroms (Å). Right: the site occupied by Mg902 in 3lkm should be occupied by a potassium ion instead [97]. The two water molecules (bottom right) should be replaced by ethylene glycol [97]. The $2mF_o - DF_c$ maps are shown at a contour level of +1.5 σ.

Although it cannot be excluded that the *cis* conformation was induced by crystallisation, these findings strongly suggested that Arg71–Phe72 *trans–cis* isomerisation plays a role in the discrimination between different effector proteins that hitherto was unknown.

Example 7—Peroxiredoxin active site in MD simulations

The human pathogen *Mycobacterium tuberculosis* is responsible for millions of deaths every year [131]. The bacterium gets engulfed by host macrophages, exposing it to a toxic environment of reactive oxygen species, but it can survive these hostile conditions by expressing peroxidases [132] such as the one-cysteine peroxiredoxin AhpE [133]. When AhpE scavenges reactive oxygen species, Cys45 is sulfenylated. The sulfenic acid form of Cys45 can be reduced by mycothiol or mycoredoxin-1 [134].

Palló *et al*. carried out MD simulations to study the active site in atomic detail (personal communication, 2015; Palló A, van Bergen L, Alonso M, Nilsson L, de Proft F, & Messens J. The revisited AhpE structures affect the MD simulations of the *Mycobacterium tuberculosis* one-cysteine peroxiredoxin). MD simulations are sensitive to errors in macromolecular structures. Palló *et al*. observed that simulations were not stable when PDB entry 1xxu [133] was used as a starting structure. The α-helix that contains Cys45 started to unwind during a 30-ns simulation. In contrast, simulations using the PDB_REDO structure were stable, probably because of the optimised hydrogen bond network in the active site (Fig. 10).

The improved AhpE structure model allows for mechanistic studies of the *M. tuberculosis* peroxiredoxin at atomic detail.

Example 8—Malaria drugs

Plasmodium falciparum is the parasite that causes malaria, which still ranks as one of the diseases with the highest death toll. The parasitic aspartic acid protease plasmepsin II is involved in degradation of the host cell haemoglobin [135] and is therefore an interesting drug target. The structure of plasmepsin II was thought to be determined in complex with two inhibitors rs367 and rs370 in PDB entries 1lee and 1lf2, respectively [136]. The difference between the two inhibitors is the position of the amino group that is *meta* in the benzamide in rs367 and *para* in rs370. The structures in 1lf2 and 1lee are nearly identical with an all-atom RMS deviation of just 0.32 Å.

Inspection of the electron density around the inhibitor in 1lee suggests that the amino group should be modelled as a *para*-substituent (see Fig. 11), which meant that both 1lee and 1lf2 contained the same inhibitor, likely as the result of a mix up

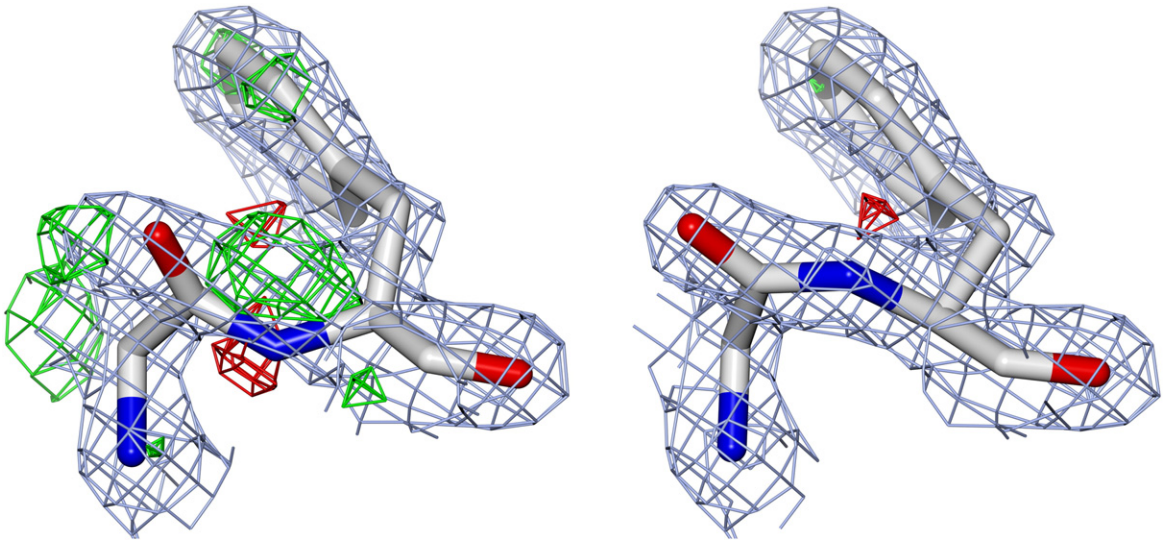


Fig. 9. Rab4a *trans–cis* flip in the switch 2 region of Rab4a. Left: the peptide between Arg71 (side chain not shown for clarity) and Phe72 in PDB entry 2bmd [130] has the *trans* conformation but deviating local geometry and the electron density around the peptide bond suggest that the peptide should have the *cis* conformation. Right: the *cis* peptide fits the experimental data much better. The 2mF_o – DF_c maps are shown at a contour level of +1.5 σ.

during the structure determination. The detection of this mix up is currently beyond the capabilities of validation routines and instead relies on critical inspection of the electron density by the crystallographer, which should be a key step in determining structures with ligands [115].

Docking studies on both structures [137] led to new candidate inhibitors, but it is a pity that twice as much time and computing power was used as needed. Low-throughput rational drug design projects aimed at better inhibitors will suffer even more from this *para–meta* error.

Example 9—The Chemistry of Autotaxin Inhibitors

Autotaxin (ATX; also known as ENPP2) is a secreted enzyme that converts lysophosphatidylcholine into the lipid signalling molecule lysophosphatidic acid. The ATX–lysophosphatidic acid signalling axis is involved in normal physiology and pathophysiology [138]. ATX expression is found to be up-regulated in several carcinomas and is implicated in motility of tumour cells [138] and, as such, a target for developing drugs for cancer

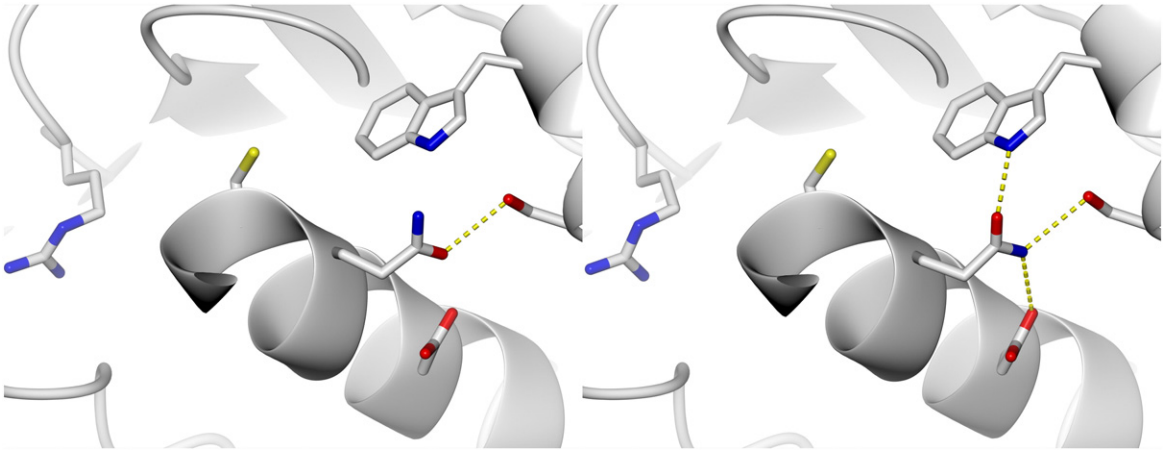


Fig. 10. Active site of one-cysteine peroxiredoxin AhpE with Cys45 in the reduced state. Left: Cys45 is located in helix α2 that unwinds during MD simulations based on PDB entry 1xxu [133]. Right: in the PDB_REDO structure, the flipped Gln46 side chain optimises the local hydrogen bonding network with Asp50, Trp80, and Ser84 and increases the stability of MD simulations.

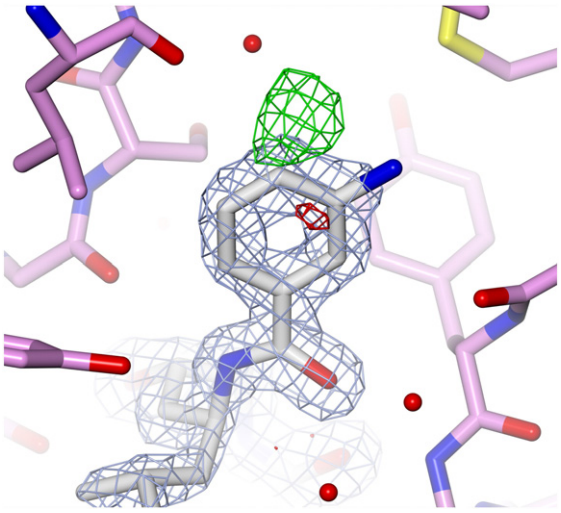


Fig. 11. The benzamide moiety of the plasmepsin II inhibitor in PDB entry 1lee [136]. The electron density suggests that the amino group of the benzamide moiety should be modelled *para* instead of *meta*. The $2mF_o - DF_c$ and $mF_o - DF_c$ maps are shown up to 1.5 Å and 2.5 Å from the displayed inhibitor fragment, respectively. The $2mF_o - DF_c$ map is shown at +2 σ.

treatment. One class of ATX inhibitors is based on a boronic acid moiety that binds covalently to the hydroxyl group of active-site residue Thr209 [139]. In this process, the hybridisation of the boron atom changes from sp^2 to sp^3, analogous to the formation of tetrahydroxyborate from boric acid.

The structure of ATX with inhibitor 3BoA (PDB entry 3wax [140]) shows the problem of dealing with changing chemistry in refinement. Although the authors correctly report that 3BoA is covalently bound to Thr209, this is not reflected in the structure model because the distance between the boron atom and the Thr-$O^{\gamma 1}$ is 2.28 Å and the boron atom is sp^2 hybridised in the model (Fig. 12). The structure of ATX and 4BoA from the same study (PDB entry 3way) suffers from the same problem. In the PDB_REDO 3wax structure model, the B-$O^{\gamma 1}$ distance is 1.38 Å and the boron atom is properly sp^3 hybridised. The earlier published structure of ATX and the boronic acid inhibitor HA155 also shows the correct geometry (PDB entry 2xrg [139]), but the other structure models may lead to misinterpretation of the ligand binding interaction.

Correct chemical representation is crucial for the design or optimisation of ATX inhibitors.

Example 10—Xylose isomerase active site

Figure 13 shows the active-site pocket of xylose isomerase, an enzyme that catalyses the interconversion between D-xylose and D-xylulose and between D-glucose and D-fructose [141]. The reaction involves hydrogen transfer and two magnesium ions.

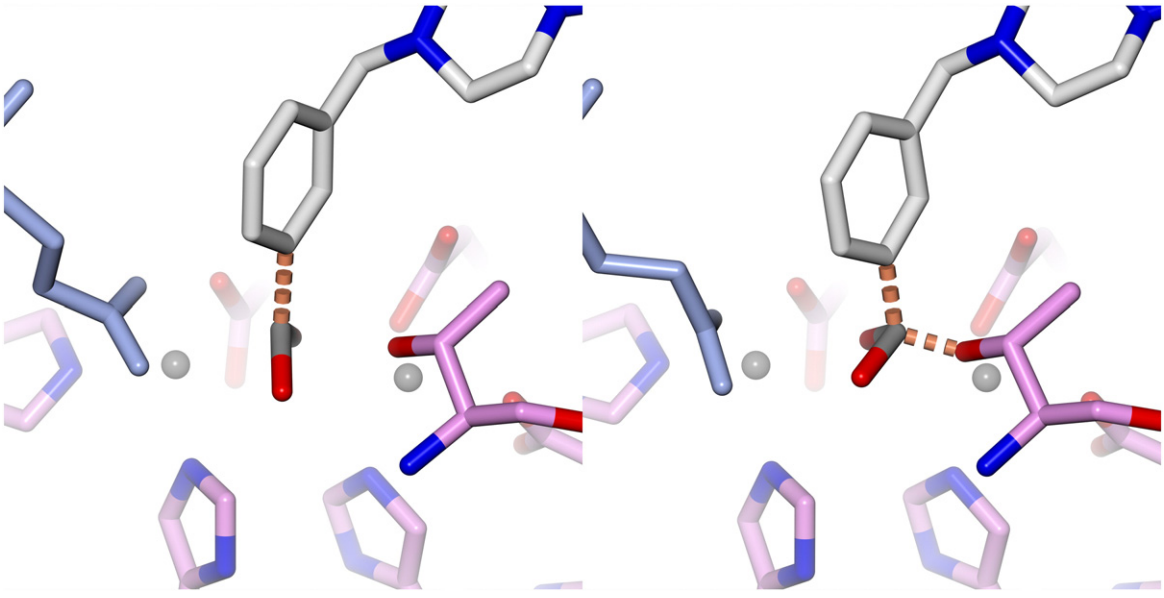


Fig. 12. Correcting chemical representation of boron. Left: fragment of the inhibitor 3BoA (white) bound to the active site of ATX (PDB entry 3wax [140], pink). The boron atom (grey) is modelled as sp^2 hybridised and covalently bound to the benzene moiety (orange broken lines) but not covalently bound to Thr209. The blue Glu576 is from a different ATX molecule related by crystal symmetry. Right: PDB_REDO-optimised version of 3wax with the correct boron hybridisation and a covalent bond to Thr209. Manually generated restraints are required to deal with the complex chemistry of the inhibitor–ATX interaction.

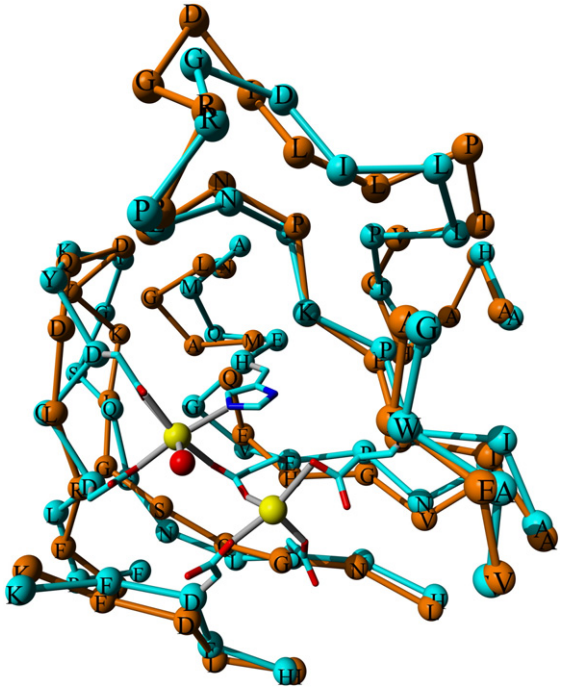


Fig. 13. Threading errors around the active-site pocket of xylose isomerase. The one-letter amino acid codes are shown on top of the C^{α} trace. Orange: 3xia [142] is mis-threaded at many locations. Cyan: 1xya [141]. Magnesium and a hydroxyl ion are shown as yellow and red spheres, respectively. This figure was prepared with YASARA [144].

PDB entry 3xia [142] was superseded by 1xya [141] after, for example, packing quality analysis [21] and inspection of the electron density revealed that many amino acids were mis-identified or mis-threaded. 1xya is in much better agreement with the biochemistry. Neutron diffraction later also showed that a water molecule rather than a hydroxyl ion should have been modelled in the active site [143].

The comparison of 3xia and 1xya clearly shows that a mis-threaded structure model can place the wrong amino acids at the wrong positions. Correct answers to biological questions related to the function of the protein can only be obtained from correctly threaded structure models. We are reasonably certain that the work of the X-ray VTF will lead to the situation that structures such as 3xia will not be passed on to the life science community. Fortunately, the deposition of reflection data is mandatory now. The problems described here might be identified more readily when experimental data are available.

Validation-related facilities

Many facilities to validate protein structures exist. Several have been mentioned in this article already. Table 2 lists a series of protein structure validation facilities that are freely accessible on the internet [112]. Of course, the validation modules mentioned here can be used not only for checking X-ray structure models but also for checking structure models derived by NMR, electron microscopy, or *in silico* modelling. Validation methods specific to NMR and electron microscopy and methods based on structure factors are beyond the scope of this article.

Much of the validation work has found its way already into the PDB_REDO project[‖]. Crystallographers can freely use the PDB_REDO server[¶] to optimise their work-in-progress structure models.

Concluding remarks

Today, experimental data deposition is an obligatory aspect of structure deposition in the PDB. Indeed, data are missing for only one recent crystal structure entry deposited in the PDB in 2014 (4ux6 [149]). We congratulate those who have set a great example by depositing datasets that were missing for 30 years (e.g., for the structural studies of leghaemoglobin by Steigemann and co-workers [150]).

Table 2. Macromolecular structure validation facilities

Facility	Description
PDB_REDO [109,145]	Constructive validation by re-refinement and partial re-building
WHAT_CHECK [43,112]	Extensive macromolecular validation
MolProbity [58,68]	Macromolecular validation
PROCHECK [41]	Protein structure geometry checks
PDB validation server [54,61]	Pre- and post-validation of PDB entries
QMEAN [62,146]	Global model-quality estimation
ProSA [23,147]	Knowledge-based potentials of mean force to evaluate macromolecular structure model accuracy
CheckMyMetal [97]	Validation of metal-binding sites
VHELIBS [82]	Validation of ligands and binding sites
ValLigURL [83]	Ligand validation
Twilight [85]	Ligand visualisation and validation
PSVS [148]	Metaserver that includes many of the above

The highly improbable features in Schwarzenbacher's structure model of the birch pollen allergen Bet v 1 protein [52] were detected [51] by studying anomalies in the statistics of PDB_REDO's model optimisation. Extremely unusual features, such as those shown in Fig. 2a, might remain undetected for longer if the corresponding reflection data are not made available.

Deposition of reflection data was not mandatory until recently (February 1, 2008). We believe that it is beneficial if missing reflection datasets are recovered and deposited because we believe that transparency by depositors and validation by others will lead to a higher-quality archive. Not necessarily because we expect any cases of fraud or gross error in PDB entries that do not have deposited reflection data but because the validation of alternative structure models against reflection data allows post-deposition model improvements. The oldest structure models in the PDB can be improved using today's methods. The average MolProbity clash score of the oldest models, for example, was at the 48th percentile relative to MolProbity's reference set and ended up at the 80th percentile after re-refinement and re-building [108]. As better computational crystallographic techniques continue to be developed, the quality of the archive can be improved ever further.

As with any scientific endeavour, validation of models and data is a component of sound application of the scientific method. Macromolecular structures solved by X-ray crystallography are the result of experiments—performed by humans—and may thus contain experimental and human errors.

Although there were a small number of individuals resistant to validation for many years [53], the majority were in favour. Research into geometric, thermodynamic, electrostatic, and many other aspects of protein structures has continued, and the series of recent highly visible cases of very unlikely structures that had to be retracted have further anchored validation tools in the protein structure solution pipelines.

Our experiences with PDB_REDO lead us to conclude that more emphasis should be placed on the deposition of raw data (diffraction images and/or unmerged reflections) and experiment-related metadata. A computer readable description of the crystallisation conditions is an important example. This will allow the development of more and better tools aimed at making the best possible structure models. We also suggest that referees of articles mentioning novel PDB entries should receive a structure validation report, without having to ask for it, rather than be made to assess a structure model's quality from the very limited information in a manuscript and its supplemental data. To this end, we reiterate the importance for crystallographers to finish and deposit their structure models and the reflection data before submitting a manuscript, rather than just before it is accepted [151]. Current PDB deposition procedures make this possible and have the option to "suppress entry titles at the time of submission to the PDB until the structure is released" [152]. In the long term, a validation report can be accompanied by a report from PDB_REDO or another automated model optimisation procedure that shows whether the model can be improved beyond the effort delivered by the depositor. This will certainly further improve the average quality of PDB structure models. These structure models, like all scientific results, are predestined to be re-used by others. Therefore, a higher average model quality will also improve the quality of many projects directed at answering important biomedical questions.

Acknowledgements

The authors thank Elmar Krieger for homology modelling advice, Rob Hooft for critically reading this manuscript, Alan Mark for pointing out to us his work on hPHMT, Anna Palló for showing us her AhpE results, and all structural biologists who have deposited experimental data. Jon Black and Coos Baakman provided technical support. G.V. acknowledges financial support from NewProt that is funded by the European Commission within its FP7 Programme, under the thematic area KBBE-2011-5 with contract number 289350, and from the research programme 11319, which is financed by S.T.W. R.P.J. is supported by Vidi 723.013.003 from Netherlands Organization for Scientific Research.

Received 16 June 2015;
Received in revised form 4 January 2016;
Accepted 1 February 2016
Available online 8 February 2016

Keywords:
structure validation;
rerefinement;
rebuilding

†http://www.nobelprize.org/nobel_prizes/chemistry/laureates/

‡http://swift.cmbi.ru.nl/

|| PDB_REDO entries are freely available from http://www.cmbi.ru.nl/pdb_redo/

¶http://xtal.nki.nl/PDB_REDO/

Abbreviations used:
CSD, Cambridge Structural Database; DACA, Directional Atomic Contact Analysis; hPNMT, human phenylethanolamine *N*-methyltransferase; MD, molecular dynamics; PBD, Polo Box Domain; PDB, Protein Data Bank; ProSA, Protein Structure Analysis; VTF, Validation Task Force.

References

[1] L. Pauling, R.B. Corey, Atomic coordinates and structure factors for two helical configurations of polypeptide chains, Proc. Natl. Acad. Sci. U. S. A. 37 (1951) 235–240, http://dx.doi.org/10.1073/pnas.37.5.235.

[2] L. Pauling, R.B. Corey, The pleated sheet, a new layer configuration of polypeptide chains, Proc. Natl. Acad. Sci. U. S. A. 37 (1951) 251–256, http://dx.doi.org/10.1073/pnas.37.5.251.

[3] J.C. Kendrew, G. Bodo, H.M. Dintzis, R.G. Parrish, H. Wyckoff, D.C. Phillips, A three-dimensional model of the myoglobin molecule obtained by X-ray analysis, Nature 181 (1958) 662–666, http://dx.doi.org/10.1038/181662a0.

[4] J.C. Kendrew, R.E. Dickerson, B.E. Strandberg, R.G. Hart, D.R. Davies, D.C. Phillips, et al., Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution, Nature 185 (1960) 422–427, http://dx.doi.org/10.1038/185422a0.

[5] M.F. Perutz, M.G. Rossmann, A.F. Cullis, H. Muirhead, G. Will, Structure of hæmoglobin: A three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis, Nature 185 (1960) 416–422, http://dx.doi.org/10.1038/185416a0.

[6] M.F. Perutz, Structure and function of haemoglobin: I. A tentative atomic model of horse oxyhaemoglobin, J. Mol. Biol. 13 (1965) 646–648.

[7] M.F. Perutz, J.C. Kendrew, H.C. Watson, Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence, J. Mol. Biol. 13 (1965) 669–678.

[8] M.G. Rossmann, The beginnings of structural biology. recollections, special section in honor of Max Perutz, Protein Sci. 3 (1994) 1731–1733, http://dx.doi.org/10.1002/pro.5560031012.

[9] W.J. Browne, North AC, Phillips DC, K. Brew, Vanaman TC, Hill RL, A possible three-dimensional structure of bovine α-lactalbumin based on that of hen's egg-white lysozyme, J. Mol. Biol. 42 (1969) 65–86, http://dx.doi.org/10.1016/0022-2836(69)90487-2.

[10] J. Novotný, R. Bruccoleri, M. Karplus, An analysis of incorrectly folded protein models. implications for structure predictions, J. Mol. Biol. 177 (1984) 787–818, http://dx.doi.org/10.1016/0022-2836(84)90049-4.

[11] J. Novotný, A.A. Rashin, R.E. Bruccoleri, Criteria that discriminate between native proteins and incorrectly folded models, Proteins 4 (1988) 19–30.

[12] D. Eisenberg, A.D. McLachlan, Solvation energy in protein folding and binding, Nature 319 (1986) 199–203, http://dx.doi.org/10.1038/319199a0.

[13] M.H. Zehfus, G.D. Rose, Compact units in proteins, Biochemistry 25 (1986) 5759–5765.

[14] S.H. Bryant, L.M. Amzel, Correctly folded proteins make twice as many hydrophobic contacts, Int. J. Pept. Protein Res. 29 (1987) 46–52, http://dx.doi.org/10.1111/j.1399-3011.1987.tb02228.x.

[15] G. Baumann, C. Frömmel, C. Sander, Polarity as a criterion in protein design, Protein Eng. 2 (1989) 329–334, http://dx.doi.org/10.1093/protein/2.5.329.

[16] M. Hendlich, P. Lackner, S. Weitckus, H. Floeckner, R. Froschauer, K. Gottsbacher, et al., Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force, J. Mol. Biol. 216 (1990) 167–180, http://dx.doi.org/10.1016/S0022-2836(05)80068-3.

[17] K. Toma, Number of residues in a sphere around a certain residue can be used as a hydrophobic penalty function of proteins, J. Mol. Graph. 9 (1991) 78–84, http://dx.doi.org/10.1016/0263-7855(91)85002-G.

[18] L. Holm, C. Sander, Evaluation of protein models by atomic solvation preference, J. Mol. Biol. 225 (1992) 93–105, http://dx.doi.org/10.1016/0022-2836(92)91028-N.

[19] R. Lüthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, Nature 356 (1992) 83–85, http://dx.doi.org/10.1038/356083a0.

[20] S.H. Bryant, C.E. Lawrence, An empirical energy function for threading protein sequence through the folding motif, Proteins 16 (1993) 92–112, http://dx.doi.org/10.1002/prot.340160110.

[21] G. Vriend, C. Sander, Quality control of protein models: Directional atomic contact analysis, J. Appl. Crystallogr. 26 (1993) 47–60, http://dx.doi.org/10.1107/S0021889892008240.

[22] C. Colovos, Yeates TO, Verification of protein structures: Patterns of nonbonded atomic interactions, Protein Sci. 2 (1993) 1511–1519, http://dx.doi.org/10.1002/pro.5560020916.

[23] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins: Struct., Funct., Bioinf. 17 (1993) 355–362, http://dx.doi.org/10.1002/prot.340170404.

[24] M. Delarue, P. Koehl, Atomic environment energies in proteins defined from statistics of accessible and contact surface areas, J. Mol. Biol. 249 (1995) 675–690, http://dx.doi.org/10.1006/jmbi.1995.0328.

[25] C.D. Stout, Crystal structures of oxidized and reduced *Azotobacter vinelandii* ferredoxin at pH 8 and 6, J. Biol. Chem. 268 (1993) 25920–25927.

[26] G.H. Stout, S. Turley, L.C. Sieker, L.H. Jensen, Structure of ferredoxin I from *Azotobacter vinelandii*, Proc. Natl. Acad. Sci. 85 (1988) 1020–1022.

[27] D. Ghosh, S. O'Donnell, W. Furey, A.H. Robbins, C.D. Stout, Iron–sulfur clusters and protein structure of *Azotobacter* ferredoxin at 2.0 Å resolution, J. Mol. Biol. 158 (1982) 73–109, http://dx.doi.org/10.1016/0022-2836(82)90451-X.

[28] E.F. Pai, U. Krengel, G.A. Petsko, R.S. Goody, W. Kabsch, A. Wittinghofer, Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: Implications for the mechanism of GTP hydrolysis, EMBO J. 9 (1990) 2351–2359.

[29] E.F. Pai, W. Kabsch, U. Krengel, K.C. Holmes, J. John, A. Wittinghofer, Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation, Nature 341 (1989) 209–214, http://dx.doi.org/10.1038/341209a0.

[30] A.M. de Vos, L. Tong, M.V. Milburn, P.M. Matias, J. Jancarik, S. Noguchi, et al., Three-dimensional structure of an oncogene protein: Catalytic domain of human c-H-ras p21, Science 239 (1988) 888–893, http://dx.doi.org/10.1126/science.2448879.

[31] T.C. Taylor, I. Andersson, The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate, J. Mol. Biol. 265 (1997) 432–444, http://dx.doi.org/10.1006/jmbi.1996.0738.

[32] S. KNIGHT, I. ANDERSSON, B.R.A.N.D.E.N. C-I, Reexamination of the three-dimensional structure of the small

subunit of RuBisCo from higher plants, Science 244 (80-) (1989) 702–705, http://dx.doi.org/10.1126/science.244.4905.702.
[33] M.S. Chapman, S.W. Suh, P.M. Curmi, D. Cascio, W.W. Smith, D.S. Eisenberg, Tertiary structure of plant RuBisCO: Domains and their contacts, Science 241 (1988) 71–74, http://dx.doi.org/10.1126/science.3133767.
[34] L. Lebioda, B. Stec, Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor, Nature 333 (1988) 683–686, http://dx.doi.org/10.1038/333683a0.
[35] L. Lebioda, B. Stec, J.M. Brewer, The structure of yeast enolase at 2.25-Å resolution. An 8-fold beta + alpha-barrel with a novel beta beta alpha alpha (beta alpha)6 topology, J. Biol. Chem. 264 (1989) 3685–3693.
[36] M.A. Navia, P.M. Fitzgerald, B.M. McKeever, C.T. Leu, J.C. Heimbach, W.K. Herber, et al., Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, Nature 337 (1989) 615–620, http://dx.doi.org/10.1038/337615a0.
[37] A. Wlodawer, M. Miller, M. Jaskólski, B.K. Sathyanarayana, E. Baldwin, I.T. Weber, et al., Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease, Science 245 (1989) 616–621, http://dx.doi.org/10.1126/science.2548279.
[38] S. McNicholas, E. Potterton, K.S. Wilson, M.E.M. Noble, Presenting your structures: The CCP4mg molecular-graphics software, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 386–394, http://dx.doi.org/10.1107/S0907444911007281.
[39] C.-I. Brändén, T.A. Jones, Between objectivity and subjectivity, Nature 343 (1990) 687–689, http://dx.doi.org/10.1038/343687a0.
[40] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr., Sect. A: Found. Crystallogr. 47 (1991) 110–119, http://dx.doi.org/10.1107/S0108767390010224.
[41] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: A program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291, http://dx.doi.org/10.1107/S0021889892009944.
[42] G. Vriend, WHAT IF: A molecular modeling and drug design program, J. Mol. Graph. 8 (1990) 52–56, http://dx.doi.org/10.1016/0263-7855(90)80070-V.
[43] R.W.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature 381 (1996) 272, http://dx.doi.org/10.1038/381272a0.
[44] A.L. Morris, M.W. MacArthur, E.G. Hutchinson, J.M. Thornton, Stereochemical quality of protein structure coordinates, Proteins Struct. Funct. Genet. 12 (1992) 345–364, http://dx.doi.org/10.1002/prot.340120407.
[45] R.A. Engh, R. Huber, Accurate bond and angle parameters for X-ray protein structure refinement, Acta Crystallogr., Sect. A: Found. Crystallogr. 47 (1991) 392–400, http://dx.doi.org/10.1107/S0108767391001071.
[46] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1963) 95–99, http://dx.doi.org/10.1016/S0022-2836(63)80023-6.
[47] M.J. Sippl, Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins, J. Mol. Biol. 213 (1990) 859–883, http://dx.doi.org/10.1016/S0022-2836(05)80269-4.
[48] J.U. Bowie, R. Lüthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science 253 (80-) (1991) 164–170, http://dx.doi.org/10.1126/science.1853201.
[49] B.J.C. Janssen, R.J. Read, A.T. Brünger, P. Gros, Crystallography: Crystallographic evidence for deviating C3b structure, Nature 448 (2007) E1–E2, http://dx.doi.org/10.1038/nature06103 (discussion E2–3).
[50] B. Borrell, Fraud rocks protein community, Nature 462 (2009) 970, http://dx.doi.org/10.1038/462970a.
[51] B. Rupp, Detection and analysis of unusual features in the structural model and structure-factor data of a birch pollen allergen, Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun. 68 (2012) 366–376, http://dx.doi.org/10.1107/S1744309112008421.
[52] N. Zaborsky, M. Brunner, M. Wallner, M. Himly, T. Karl, R. Schwarzenbacher, et al., Antigen aggregation decides the fate of the allergic immune response, J. Immunol. 184 (2010) 725–735, http://dx.doi.org/10.4049/jimmunol.0902080.
[53] G.A. Petsko, Large cast, but no plot, Nature 359 (1992) 596–597, http://dx.doi.org/10.1038/359596a0.
[54] R.J. Read, P.D. Adams, W.B. Arendall, A.T. Brunger, P. Emsley, R.P. Joosten, et al., A new generation of crystallographic validation tools for the Protein Data Bank, Structure 19 (2011) 1395–1412, http://dx.doi.org/10.1016/j.str.2011.08.006.
[55] G.T. Montelione, M. Nilges, A. Bax, P. Güntert, T. Herrmann, J.S. Richardson, et al., Recommendations of the wwPDB NMR Validation Task Force, Structure 21 (2013) 1563–1570, http://dx.doi.org/10.1016/j.str.2013.07.021.
[56] R. Henderson, A. Sali, M.L. Baker, B. Carragher, B. Devkota, K.H. Downing, et al., Outcome of the first electron microscopy Validation Task Force meeting, Structure 20 (2012) 205–214, http://dx.doi.org/10.1016/j.str.2011.12.014.
[57] H. Berman, K. Henrick, H. Nakamura, J.L. Markley, The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data, Nucleic Acids Res. 35 (2007) D301–D303, http://dx.doi.org/10.1093/nar/gkl971.
[58] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography, Acta Crystallogr. Sect. D: Biol. Crystallogr. 66 (2010) 12–21, http://dx.doi.org/10.1107/S0907444909042073.
[59] J.F. Doreleijers, A.W. Sousa da Silva, E. Krieger, S.B. Nabuurs, C.A.E.M. Spronk, T.J. Stevens, et al., CING: An integrated residue-based structure validation program suite, J. Biomol. NMR 54 (2012) 267–283, http://dx.doi.org/10.1007/s10858-012-9669-7.
[60] A. Gutmanas, Y. Alhroub, G.M. Battle, J.M. Berrisford, E. Bochet, M.J. Conroy, et al., PDBe: Protein Data Bank in Europe, Nucleic Acids Res. 42 (2014) D285–D291, http://dx.doi.org/10.1093/nar/gkt1180.
[61] S. Gore, S. Velankar, G.J. Kleywegt, Implementing an X-ray validation pipeline for the Protein Data Bank, Acta Crystallogr. Sect. D: Biol. Crystallogr. 68 (2012) 478–483, http://dx.doi.org/10.1107/S0907444911050359.
[62] P. Benkert, S.C.E. Tosatto, D. Schomburg, QMEAN: A comprehensive scoring function for model quality assessment, Proteins 71 (2008) 261–277, http://dx.doi.org/10.1002/prot.21715.
[63] J. Haas, S. Roth, K. Arnold, F. Kiefer, T. Schmidt, L. Bordoli, et al., The protein model portal—A comprehensive resource for protein structure and model information, Database 2013 (2013), http://dx.doi.org/10.1093/database/bat031.

[64] C.R. Groom, F.H. Allen, The Cambridge Structural Database in retrospect and prospect, Angew. Chem. Int. Ed. 53 (2014) 662–671, http://dx.doi.org/10.1002/anie.201306438.
[65] R.A. Engh, R. Huber, in: Rossmann MG, E. Arnold (Eds.), Structure Quality and Target Parameters, vol. F. 1st ed.International Union of Crystallography, Chester, England 2001, pp. 382–392.
[66] R.W.W. Hooft, C. Sander, G. Vriend, Verification of protein structures: Side-chain planarity, J. Appl. Crystallogr. 29 (1996) 714–716, http://dx.doi.org/10.1107/S0021889896008631.
[67] R.W.W. Hooft, C. Sander, G. Vriend, Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures, Proteins Struct. Funct. Genet. 26 (1996) 363–376, http://dx.doi.org/10.1002/(SICI)1097-0134(199612)26:4<363::AID-PROT1>3.0.CO;2-D.
[68] I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, et al., MolProbity: All-atom contacts and structure validation for proteins and nucleic acids, Nucleic Acids Res. 35 (2007) W375–W383, http://dx.doi.org/10.1093/nar/gkm216.
[69] J.E. Nielsen, K.V. Andersen, B. Honig, R.W. Hooft, G. Klebe, G. Vriend, et al., Improving macromolecular electrostatics calculations, Protein Eng. 12 (1999) 657–662, http://dx.doi.org/10.1093/protein/12.8.657.
[70] J.E. Nielsen, G. Vriend, Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pK_a calculations, Proteins Struct. Funct. Genet. 43 (2001) 403–412, http://dx.doi.org/10.1002/prot.1053.
[71] G. Vriend, R. Hooft, Some WHAT_CHECK checks explained, Protein Data Bank Q. Newsl. 84 (1998) 4–5.
[72] EU 3-D Validation Network, Who checks the checkers? Four validation tools applied to eight atomic resolution structures, J. Mol. Biol. 276 (1998) 417–436, http://dx.doi.org/10.1006/jmbi.1997.1526.
[73] R.W.W. Hooft, C. Sander, G. Vriend, Objectively judging the quality of a protein structure from a Ramachandran plot, Comput. Appl. Biosci. 13 (1997) 425–430, http://dx.doi.org/10.1093/bioinformatics/13.4.425.
[74] D.S. Berkholz, M.V. Shapovalov, Dunbrack RL Jr., Karplus PA, Conformation dependence of backbone geometry in proteins, Structure 17 (2009) 1316–1325, http://dx.doi.org/10.1016/j.str.2009.08.012.
[75] W.G. Touw, G. Vriend, On the complexity of Engh and Huber refinement restraints: The angle τ as example, Acta Crystallogr. Sect. D: Biol. Crystallogr. 66 (2010) 1341–1350, http://dx.doi.org/10.1107/S0907444910040928.
[76] V. De Filippis, C. Sander, G. Vriend, Predicting local structural changes that result from point mutations, Protein Eng. 7 (1994) 1203–1208.
[77] R.L. Dunbrack, F.E. Cohen, Bayesian statistical analysis of protein side-chain rotamer preferences, Protein Sci. 6 (1997) 1661–1681, http://dx.doi.org/10.1002/pro.5560060807.
[78] S.C. Lovell, J.M. Word, J.S. Richardson, D.C. Richardson, The penultimate rotamer library, Proteins 40 (2000) 389–408.
[79] K.R.M. Berntsen, G. Vriend, Anomalies in the refinement of isoleucine, Acta Crystallogr. Sect. D: Biol. Crystallogr. 70 (2014) 1037–1049, http://dx.doi.org/10.1107/S139900471400087X.
[80] S.C. Lovell, I.W. Davis, W.B. Arendall, P.I.W. De Bakker, J.M. Word, M.G. Prisant, et al., Structure validation by C^{α} geometry: ϕ,ψ and C^{β} deviation, Proteins Struct. Funct. Genet. 50 (2003) 437–450, http://dx.doi.org/10.1002/prot.10286.
[81] W. Sheffler, D. Baker, RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design, and validation, Protein Sci. 18 (2009) 229–239, http://dx.doi.org/10.1002/pro.8.
[82] A. Cereto-Massagué, M.J. Ojeda, R.P. Joosten, C. Valls, M. Mulero, M.J. Salvado, et al., The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites, J. Cheminf. 5 (2013) 36, http://dx.doi.org/10.1186/1758-2946-5-36.
[83] G.J. Kleywegt, M.R. Harris, ValLigURL: A server for ligand-structure comparison and validation, Acta Crystallogr. Sect. D: Biol. Crystallogr. 63 (2007) 935–938, http://dx.doi.org/10.1107/S090744490703315X.
[84] I.J. Bruno, J.C. Cole, M. Kessler, J. Luo, W.D.S. Momerwell, L.H. Purkis, et al., Retrieval of crystallographically-derived molecular geometry information, J. Chem. Inf. Comput. Sci. 44 (2004) 2133–2144, http://dx.doi.org/10.1021/ci049780b.
[85] C.X. Weichenberger, E. Pozharski, B. Rupp, Visualizing ligand molecules in twilight electron density, Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun. 69 (2013) 1–6, http://dx.doi.org/10.1107/S1744309112044387.
[86] T. Lütteke, C.-W. von der Lieth, pdb-care (PDB carbohydrate residue check): A program to support annotation of complex carbohydrate structures in PDB files, BMC Bioinf. 5 (2004) 69, http://dx.doi.org/10.1186/1471-2105-5-69.
[87] T. Lütteke, M. Frank, C.W. von der Lieth, Carbohydrate structure suite (CSS): Analysis of carbohydrate 3D structures derived from the PDB, Nucleic Acids Res. 33 (2005), http://dx.doi.org/10.1093/nar/gki013.
[88] J. Agirre, J. Iglesias-Fernández, C. Rovira, G.J. Davies, K.S. Wilson, K.D. Cowtan, Privateer: software for the conformational validation of carbohydrate structures, Nat. Struct. Mol. Biol. 22 (2015) 833–834, http://dx.doi.org/10.1038/nsmb.3115.
[89] J. Agirre, G. Davies, K. Wilson, K. Cowtan, Carbohydrate anomalies in the PDB, Nat. Chem. Biol. 11 (2015) 303, http://dx.doi.org/10.1038/nchembio.1798.
[90] T. Lütteke, Analysis and validation of carbohydrate three-dimensional structures, Acta Crystallogr. Sect. D: Biol. Crystallogr. 65 (2009) 156–168, http://dx.doi.org/10.1107/S0907444909001905.
[91] P. Emsley, A.T. Brunger, T. Lütteke, Tools to Assist Determination and Validation of Carbohydrate 3D Structure Data, 2015 229–240, http://dx.doi.org/10.1007/978–1-4939-2343-4_17.
[92] F.-C. Chou, P. Sripakdeevong, S.M. Dibrov, T. Hermann, R. Das, Correcting pervasive errors in RNA crystallography through enumerative structure prediction, Nat. Methods 10 (2013) 74–76, http://dx.doi.org/10.1038/nmeth.2262.
[93] M. Nayal, E. Di Cera, Valence screening of water in protein crystals reveals potential Na^+ binding sites, J. Mol. Biol. 256 (1996) 228–234, http://dx.doi.org/10.1006/jmbi.1996.0081.
[94] I.D. Brown, Predicting bond lengths in inorganic crystals, Acta Crystallogr. Sect. B: Struct. Crystallogr. Cryst. Chem. 33 (1977) 1305–1310, http://dx.doi.org/10.1107/S0567740877005998.
[95] I.D. Brown, Chemical and steric constraints in inorganic solids, Acta Crystallogr. Sect. B: Struct. Sci. 48 (1992) 553–572, http://dx.doi.org/10.1107/S0108768192002453.
[96] P. Müller, S. Köpke, G.M. Sheldrick, Is the bond-valence method able to identify metal atoms in protein structures? Acta Crystallogr. Sect. D: Biol. Crystallogr. 59 (2002) 32–37, http://dx.doi.org/10.1107/S0907444902018000.
[97] H. Zheng, M.D. Chordia, D.R. Cooper, M. Chruszcz, P. Müller, G.M. Sheldrick, et al., Validation of metal-binding sites in macromolecular structures with the CheckMyMetal

Web server, Nat. Protoc. 9 (2014) 156–170, http://dx.doi.org/10.1038/nprot.2013.172.
[98] N. Echols, N. Morshed, P.V. Afonine, A.J. McCoy, M.D. Miller, R.J. Read, et al., Automated identification of elemental ions in macromolecular crystal structures, Acta Crystallogr. Sect. D: Biol. Crystallogr. 70 (2014) 1104–1114, http://dx.doi.org/10.1107/S1399004714001308.
[99] Z. Dauter, A. Wlodawer, W. Minor, M. Jaskolski, B. Rupp, Avoidable errors in deposited macromolecular structures: An impediment to efficient data mining, IUCrJ 1 (2014) 1–15, http://dx.doi.org/10.1107/S2052252514005442.
[100] J.B. Thoden, S.M. Firestine, S.J. Benkovic, H.M. Holden, PurT-encoded glycinamide ribonucleotide transformylase. accommodation of adenosine nucleotide analogs within the active site, J. Biol. Chem. 277 (2002) 23898–23908, http://dx.doi.org/10.1074/jbc.M202251200.
[101] R. Huber, W. Steigemann, Two *cis*-prolines in the Bence-Jones protein Rei and the *cis*-pro-bend, FEBS Lett. 48 (1974) 2–4.
[102] D.E. Stewart, a Sarkar, J.E. Wampler, Occurrence and role of *cis* peptide bonds in protein structures, J. Mol. Biol. 214 (1990) 253–260, http://dx.doi.org/10.1016/0022-2836(90)90159-J.
[103] M.S. Weiss, A. Jabs, R. Hilgenfeld, Peptide bonds revisited, Nat. Struct. Biol. 5 (1998) 676.
[104] A. Jabs, M.S. Weiss, R. Hilgenfeld, A method to detect nonproline *cis* peptide bonds in proteins, J. Mol. Biol. 286 (1999) 291–304.
[105] W.G. Touw, R.P. Joosten, G. Vriend, Detection of *trans-cis* flips and peptide-plane flips in protein structures, Acta Crystallogr. Sect. D: Biol. Crystallogr. 71 (2015), http://dx.doi.org/10.1107/S1399004715008263.
[106] R.P. Joosten, G. Vriend, PDB improvement starts with data deposition, Science 317 (2007) 195–196, http://dx.doi.org/10.1126/science.317.5835.195.
[107] R.P. Joosten, J. Salzemann, V. Bloch, H. Stockinger, A.-C. Berglund, C. Blanchet, et al., PDB_REDO: Automated re-refinement of X-ray structure models in the PDB, J. Appl. Crystallogr. 42 (2009) 376–384, http://dx.doi.org/10.1107/S0021889809008784.
[108] R.P. Joosten, K. Joosten, S.X. Cohen, G. Vriend, A. Perrakis, Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank, Bioinformatics 27 (2011) 3392–3398, http://dx.doi.org/10.1093/bioinformatics/btr590.
[109] R.P. Joosten, K. Joosten, G.N. Murshudov, A. Perrakis, PDB_REDO : Constructive validation, more than just looking for errors, Acta Crystallogr. Sect. D: Biol. Crystallogr. 68 (2012) 484–496, http://dx.doi.org/10.1107/S0907444911054515.
[110] G.N. Murshudov, P. Skubák, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, et al., REFMAC5 for the refinement of macromolecular crystal structures, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 355–367, http://dx.doi.org/10.1107/S0907444911001314.
[111] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637, http://dx.doi.org/10.1002/bip.360221211.
[112] W.G. Touw, C. Baakman, J. Black, T.A.H. Beek, E. Krieger, P. Joosten, et al., A series of PDB-related databanks for everyday needs, Nucleic Acids Res. 43 (2015) D364–D368, http://dx.doi.org/10.1093/nar/gku1028.
[113] R.P. Joosten, T.A.H. te Beek, E. Krieger, M.L. Hekkelman, R.W.W. Hooft, R. Schneider, et al., A series of PDB related databases for everyday needs, Nucleic Acids Res. 39 (2011) D411–D419, http://dx.doi.org/10.1093/nar/gkq1105.
[114] G.J. Kleywegt, Crystallographic refinement of ligand complexes, Acta Crystallogr. Sect. D: Biol. Crystallogr. 63 (2006) 94–100, http://dx.doi.org/10.1107/S0907444906022657.
[115] E. Pozharski, C.X. Weichenberger, B. Rupp, Techniques, tools and best practices for ligand electron-density analysis and results from their application to deposited crystal structures, Acta Crystallogr. Sect. D: Biol. Crystallogr. 69 (2013) 150–167, http://dx.doi.org/10.1107/S0907444912044423.
[116] E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, et al., Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8, Proteins 77 (Suppl. 9) (2009) 114–122, http://dx.doi.org/10.1002/prot.22570.
[117] B. García-Alvarez, G. de Cárcer, S. Ibañez, E. Bragado-Nilsson, G. Montoya, Molecular and structural basis of polo-like kinase 1 substrate recognition: Implications in centrosomal localization, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 3107–3112, http://dx.doi.org/10.1073/pnas.0609131104.
[118] K. Strebhardt, A. Ullrich, Targeting polo-like kinase 1 for cancer therapy, Nat. Rev. Cancer 6 (2006) 321–330, http://dx.doi.org/10.1038/nrc1841.
[119] D.J. Huggins, G.J. McKenzie, D.D. Robinson, A.J. Narváez, B. Hardwick, M. Roberts-Thomson, et al., Computational analysis of phosphopeptide binding to the polo-Box domain of the mitotic kinase PLK1 using molecular dynamics simulation, PLoS Comput. Biol. 6 (2010), e1000880, http://dx.doi.org/10.1371/journal.pcbi.1000880.
[120] A.E.H. Elia, P. Rellos, L.F. Haire, J.W. Chao, F.J. Ivins, K. Hoepker, et al., The molecular basis for phosphodependent substrate targeting and regulation of Plks by the polo-box domain, Cell 115 (2003) 83–95, http://dx.doi.org/10.1016/S0092-8674(03)00725-6.
[121] W.J. Qian, J.E. Park, D. Lim, S.Y. Park, K.W. Lee, M.B. Yaffe, et al., Peptide-based inhibitors of Plk1 polo-box domain containing mono-anionic phosphothreonine esters and their pivaloyloxymethyl prodrugs, Chem. Biol. 20 (2013) 1255–1264, http://dx.doi.org/10.1016/j.chembiol.2013.09.005.
[122] N. Drinkwater, H. Vu, K.M. Lovell, K.R. Criscione, B.M. Collins, T.E. Prisinzano, et al., Fragment-based screening by X-ray crystallography, MS and isothermal titration calorimetry to identify PNMT (phenylethanolamine *N*-methyltransferase) inhibitors, Biochem. J. 431 (2010) 51–61, http://dx.doi.org/10.1042/BJ20100651.
[123] P.C. Nair, A.K. Malde, N. Drinkwater, A.E. Mark, Missing fragments: Detecting cooperative binding in fragment-based drug design, ACS Med. Chem. Lett. 3 (2012) 322–326, http://dx.doi.org/10.1021/ml300015u.
[124] P.P. Di Fiore, J.H. Pierce, M.H. Kraus, O. Segatto, C.R. King, Aaronson SA, erbB-2 is a potent oncogene when overexpressed in NIH/3T3 cells, Science 237 (1987) 178–182, http://dx.doi.org/10.1126/science.2885917.
[125] D.J. Slamon, B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, et al., Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2, vol. 3442001, http://dx.doi.org/10.1056/NEJM200103153441101.
[126] H.-S. Cho, K. Mason, K.X. Ramyar, A.M. Stanley, S.B. Gabelli, D.W. Denney, et al., Structure of the extracellular region of HER2 alone and in complex with the Herceptin fab, Nature 421 (2003) 756–760, http://dx.doi.org/10.1038/nature01392.
[127] M. Vicente-Manzanares, X. Ma, R.S. Adelstein, A.R. Horwitz, Non-muscle myosin II takes centre stage in cell

adhesion and migration, Nat. Rev. Mol. Cell Biol. 10 (2009) 778–790, http://dx.doi.org/10.1038/nrm2786.
[128] Q. Ye, S.W. Crawley, Y. Yang, G.P. Côté, Z. Jia, Crystal structure of the alpha-kinase domain of *Dictyostelium* myosin heavy chain kinase A, Sci. Signal. 3 (2010) ra17, http://dx.doi.org/10.1126/scisignal.2000525.
[129] R.S. D'Souza, R. Semus, E.A. Billings, C.B. Meyer, K. Conger, J.E. Casanova, Rab4 orchestrates a small GTPase cascade for recruitment of adaptor proteins to early endosomes, Curr. Biol. 24 (2014) 1187–1198, http://dx.doi.org/10.1016/j.cub.2014.04.003.
[130] S.K. Huber, A.J. Scheidig, High resolution crystal structures of human Rab4a in its active and inactive conformations, FEBS Lett. 579 (2005) 2821–2829, http://dx.doi.org/10.1016/j.febslet.2005.04.020.
[131] WHO, Global Tuberculosis Report 2014 (WHO/HTM/TB/2014.08), 2014 (doi:WHO/HTM/TB/2014.08).
[132] C. Manca, S. Paul, C.E. Barry, V.H. Freedman, G. Kaplan, *Mycobacterium tuberculosis* catalase and peroxidase activities and resistance to oxidative killing in human monocytes *in vitro*, Infect. Immun. 67 (1999) 74–79.
[133] S. Li, N.A. Peterson, M.Y. Kim, C.Y. Kim, L.W. Hung, M. Yu, et al., Crystal structure of AhpE from *Mycobacterium tuberculosis*, a 1-Cys peroxiredoxin, J. Mol. Biol. 346 (2005) 1035–1046, http://dx.doi.org/10.1016/j.jmb.2004.12.046.
[134] K. Van Laer, L. Buts, N. Foloppe, D. Vertommen, K. Van Belle, K. Wahni, et al., Mycoredoxin-1 is one of the missing links in the oxidative stress defence mechanism of mycobacteria, Mol. Microbiol. 86 (2012) 787–804, http://dx.doi.org/10.1111/mmi.12030.
[135] S. Le Bonniec, C. Deregnaucourt, V. Redeker, R. Banerjee, P. Grellier, D.E. Goldberg, et al., Plasmepsin II, an acidic hemoglobinase from the *Plasmodium falciparum* food vacuole, is active at neutral pH on the host erythrocyte membrane skeleton, J. Biol. Chem. 274 (1999) 14218–14223, http://dx.doi.org/10.1074/jbc.274.20.14218.
[136] O.A. Asojo, E. Afonina, S.V. Gulnik, B. Yu, J.W. Erickson, R. Randad, et al., Structures of Ser205 mutant plasmepsin II from *Plasmodium falciparum* at 1.8 Å in complex with the inhibitors rs367 and rs370, Acta Crystallogr. Sect. D: Biol. Crystallogr. 58 (2002) 2001–2008, http://dx.doi.org/10.1107/S0907444902014695.
[137] V. Kasam, M. Zimmermann, A. Maaß, H. Schwichtenberg, A. Wolf, N. Jacq, et al., Design of new plasmepsin inhibitors: A virtual high throughput screening approach on the EGEE grid, J. Chem. Inf. Model. 47 (2007) 1818–1828, http://dx.doi.org/10.1021/ci600451t.
[138] W.H. Moolenaar, A. Perrakis, Insights into autotaxin: How to produce and present a lipid mediator, Nat. Rev. Mol. Cell Biol. 12 (2011) 674–679, http://dx.doi.org/10.1038/nrm3188.
[139] J. Hausmann, S. Kamtekar, E. Christodoulou, J.E. Day, T. Wu, Z. Fulkerson, et al., Structural basis of substrate discrimination and integrin binding by autotaxin, Nat. Struct. Mol. Biol. 18 (2011) 198–204, http://dx.doi.org/10.1038/nsmb.1980.
[140] M. Kawaguchi, T. Okabe, S. Okudaira, H. Nishimasu, R. Ishitani, H. Kojima, et al., Screening and X-ray crystal structure-based optimization of autotaxin (ENPP2) inhibitors, using a newly developed fluorescence probe, ACS Chem. Biol. 8 (2013) 1713–1721, http://dx.doi.org/10.1021/cb400150c.
[141] A. Lavie, K.N. Allen, G.A. Petsko, D. Ringe, X-ray crystallographic structures of D-xylose isomerase-substrate complexes position the substrate and provide evidence for metal movement during catalysis, Biochemistry 33 (1994) 5469–5480, http://dx.doi.org/10.1021/bi00184a016.
[142] G.K. Farber, A. Glasfeld, G. Tiraby, D. Ringe, G.A. Petsko, Crystallographic studies of the mechanism of xylose isomerase, Biochemistry 28 (1989) 7289–7297, http://dx.doi.org/10.1021/bi00444a022.
[143] A.K. Katz, X. Li, H.L. Carrell, B.L. Hanson, P. Langan, L. Coates, et al., Locating active-site hydrogen atoms in D-xylose isomerase: Time-of-flight neutron diffraction, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 8342–8347, http://dx.doi.org/10.1073/pnas.0602598103.
[144] E. Krieger, G. Vriend, YASARA view—Molecular graphics for all devices-from smartphones to workstations, Bioinformatics 1–2 (2014), http://dx.doi.org/10.1093/bioinformatics/btu426.
[145] R.P. Joosten, F. Long, G.N. Murshudov, A. Perrakis, The PDB_REDO server for macromolecular structure model optimization, IUCrJ 1 (2014) 213–220, http://dx.doi.org/10.1107/S2052252514009324.
[146] P. Benkert, M. Künzli, T. Schwede, QMEAN server for protein model quality estimation, Nucleic Acids Res. 37 (2009), http://dx.doi.org/10.1093/nar/gkp322.
[147] M. Wiederstein, M.J. Sippl, ProSA-web: Interactive Web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007), http://dx.doi.org/10.1093/nar/gkm290.
[148] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by Structural Genomics Consortia, Proteins Struct. Funct. Genet. 66 (2007) 778–795, http://dx.doi.org/10.1002/prot.21165.
[149] D.R. Cheshire, A. Åberg, G.M.K. Andersson, G. Andrews, H.G. Beaton, T.N. Birkinshaw, et al., The discovery of novel, potent and highly selective inhibitors of inducible nitric oxide synthase (iNOS), Bioorg. Med. Chem. Lett. 21 (2011) 2468–2471, http://dx.doi.org/10.1016/j.bmcl.2011.02.061.
[150] E.G. Arutynyan, P. Kuranova, B.K. Vainshtein, W. Steigemann, X-ray structural investigation of leghemoglobin. VI. Structure of acetate-ferrileghemoglobin at a resolution of 2.0 Å, Kristallografiya 25 (1980) 80–103.
[151] R.P. Joosten, H. Soueidan, L.F.A. Wessels, A. Perrakis, Timely deposition of macromolecular structures is necessary for peer review, Acta Crystallogr. Sect. D: Biol. Crystallogr. 69 (2013) 2293–2295, http://dx.doi.org/10.1107/S0907444913024621.
[152] H. Berman, G.J. Kleywegt, H. Nakamura, Markley JL, Comment on timely deposition of macromolecular structures is necessary for peer review by Joosten et al. (2013), Acta Crystallogr. Sect. D: Biol. Crystallogr. 69 (2013) 2296, http://dx.doi.org/10.1107/S0907444913029168.